Active Nonrigid ICP Algorithm

Shiyang Cheng[1], Ioannis Marras[1], Stefanos Zafeiriou[1] and Maja Pantic[1,2]

[1] Department of Computing, Imperial College London, United Kingdom

[2] EEMCS, University of Twente, Netherlands

{shiyang.cheng11, i.marras, s.zafeiriou, m.pantic}@imperial.ac.uk

Abstract— The problem of fitting a 3D facial model to a 3D mesh has received a lot of attention the past 15-20 years. The majority of the techniques fit a general model consisting of a simple parameterisable surface or a mean 3D facial shape. The drawback of this approach is that is rather difficult to describe the non-rigid aspect of the face using just a single facial model. One way to capture the 3D facial deformations is by means of a statistical 3D model of the face or its parts. This is particularly evident when we want to capture the deformations of the mouth region. Even though statistical models of face are generally applied for modelling facial intensity, there are few approaches that fit a statistical model of 3D faces. In this paper, in order to capture and describe the non-rigid nature of facial surfaces we build a part-based statistical model of the 3D facial surface and we combine it with non-rigid iterative closest point algorithms. We show that the proposed algorithm largely outperforms state-of-the-art algorithms for 3D face fitting and alignment especially when it comes to the description of the mouth region.

I. INTRODUCTION

Building and fitting of 3D facial models is a field that has received a lot of attention in the past decade [1], [2], [3], [4], [5], [6], [7], [8]. 3D fitting algorithms are very important, since they can be used to identify particular facial landmarks or to recognize and define facial deformations that are important for analysis of facial expressions. This research field received further attention recently with the introduction of low cost depth cameras, such as Kinect. Even though, the majority of the methodologies use just a simple pre-defined mesh model, parameterized or not, to fit a test mesh [9], [3], [5], [8], [10]. Unfortunately, such methodologies may fail to describe properly the complex, non-linear and highly deformable structure of the face.

In this paper, we examine the problem of fitting a 3D facial models to high-resolution depth scan. Our key contributions are summarized as follows:

- To provide a point-driven mesh deformation procedure that takes the advantage of robust 2D face alignment [11] and deforms the 3D template model. This pre-processing step helps to prevent incorrect fitting of face part.
- To propose a dynamic local fitting procedure that makes full use of dynamic subdivision framework. To this end, we successfully adopt the non-rigid Iterative Closest Point (ICP) algorithm [5] to this procedure, and show that it manages to accurately model the subtle facial feature.
- To present a new active method for describing and fitting 3D faces by learning a local statistical model

of facial parts, and combining them with the non-rigid ICP [5]. This method is also adopted to our dynamic local fitting procedure.

II. RELATED WORK

Various 3D face fitting methods have been proposed to address specific problems in different scenarios, such as the situation with high-resolution data [3], partial range scan [12], [5] and normal maps [13]. Passalis et al. [3] proposed a deformable fitting framework based on AFM (Annotated Face Model) [9], [14], which has been one of the state-of-the-art 3D fitting methods. It employs a physics-based model, and defines mass, damping and stiffness matrix to control the deformation of template. To deform the template, an external force that is proportional to the euclidean distance between limit surface and closest points in target mesh is defined. The resulting cost function is solved using Finite Element Method (FEM) approximation. In order to increase flexibility and scalability subdivision surfaces [15] were used in [3]. The model was further used to describe 3D faces in various poses by exploiting the fact that faces are approximate symmetric [16]. The model was particular effective for capturing faces for the purposes of face recognition and verification, due to the fact that fitting relayed of deforming a single facial template, in many cases, it was unable to capture facial deformations due to facial expressions [17]. In [17] in order to better capture facial expressions an extra fitting strategy using Thin Plate Splines (TPS) was applied.

Another family of 3D fitting algorithms that use a single template is non-rigid ICP (Iterative Closest Point), where correspondence of points are found by a search based on spatial proximity, and the transformation of each point is modelled by general deformation. The main challenge of non-rigid ICP is how to tackle the high-dimensional optimization problem for local deformations while preserving the convergence properties. Hahnel et al. [18] propose a hierarchical method that firstly fits a skeleton structure to the scan points and then aligns local parts. Allen et al. [6] assume an affine transformation at each point of the template mesh to allow non-rigid registrations of full-body scans to a high-resolution template. Similar to this approach, the optimal non-rigid ICP (NICP) step proposed by Amberg et al. [5] assumes local affine transformation for each vertex, additionally it defines a series of stiffness weights to regularize the deformation. NICP has demonstrated fast convergence and reliable fitting on a number of examples. However, since the NICP optimization is essentially solving a linear system, it

remains as a question whether the system is solvable or not for very dense template.

Owing to the representative power and generalization ability of statistical shape models [1], [2], various works have successfully employed statistical shape prior in non-rigid registration [19], [8], [20], [21], [10], [7]. Albrecht et al. [19] proposed to use Tikhonov regularization that is based on a statistical deformation model to regularize the shape. Schneider et al. [8] used a morphable model to control the deformation, and jointly optimized the non-rigid parameters with the rigid pose in a unified ICP framework. [8] is further extended to the problem of head scan registration [20]. Similar to [8], [20], Amberg et al. [10] fitted the 3DMM to the weighted correspondences that are retrieved by ICP algorithm. Pan et al. [7] further defined the deformation offsets to each vertex of the shape, and solved the offsets and non-rigid parameters alternately.

Statistical shape model has seldom been introduced to the non-rigid ICP framework. Although NICP [5] is able to provide closely fitted surface, it has weak constraint on the shape geometry, which might lead to noisy fitting result and even non-face like fittings. Here, we propose to incorporate statistical prior into the fitting procedure of NICP [5]. The use of subspace essentially puts extra constraint on the fitting procedure that preserves the structure of each face part. Besides, the fitting directed by the linear face model is robust to the noisy raw scan, leading to a smoother and finer representation of the target surface. We integrate the proposed fitting into a dynamic subdivision framework, making it possible to accurately capture the subtle details in a high-resolution facial scans. In addition, we provide a point-driven mesh deformation procedure that generates deformed reference shape based on robust 2D face alignment.

III. DYNAMIC SUBDIVISION FRAMEWORK

The core idea of our framework is to dynamically fit facial data using a deformable 3D face model, and provide an accurately fitted surface. In contrary to previous works on 3D surface registration [3], [9], [14] that subsample the data using an annotated template to gain efficiency, our method starts from a sparse level and dynamically propagates to subsequent levels, in which the fittings are performed locally to model regional deformation. We argue that the subsampling step sacrifices the accuracy and loses essential details (for example, wrinkles in the forehead) for expression recognition task. Our framework mainly consists of three processing stages: (1) Point-driven template deformation; (2) Rigid alignment; (3) Local deformable model fitting.

A. Point-driven Template Deformation

Our fitting starts from annotated model of the face (AFM) [14], [9], which is segmented into several parts according to the facial physiology information [14], [9]. For the purpose of local fitting, we further manually segment the subdivided AFM (see Fig. 1 for details).

A natural problem of using AFM is that, for the exaggerated expression like surprise, final annotated parts tend to
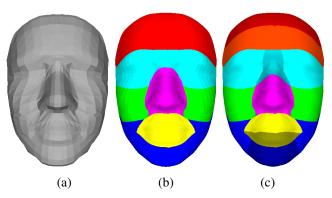

(a)　　　　　(b)　　　　　(c)

Fig. 1: Annotated areas for different levels of AFM. (a) The starting level, without subdivision and segmentation; (b) First subdivision level, with 6 annotated face parts (cheeks are combined into one part); (c) Second subdivision level, with 13 individual parts.

lose original physiology meaning. Specifically, the detected mouth frequently lies in between nose and upper lip. This is because the initial position of mouth hardly covers the entire region, leading to a false preliminary correspondence retrieved by closest points matching. Even though with predefined expression-wise AFMs, this problem can hardly be solved, since the valence and arousal of emotions vary from time to time, thus expression-wise templates simply cannot cover them.

Point-driven mesh deformation: To tackle this problem, we propose to deform AFM based on the fitting of 2D face alignment. Considering the robustness and efficiency of 2D face alignment [11], [22], [23], it is reliable to deform the mesh under the guidance of 2D fitting. Therefore, we use publicly available Chehra face tracker [23], and perform fitting on the color image of 3D scan.

To deform a 3D model, we retrieve the corresponding 3D coordinates of 2D fitting in the raw scan. Based on the 2D annotation scheme, we annotated 33 landmarks as the anchor points on the AFM (Fig. 2), note that we excluded all boundary points since they are generally noisy. Besides, we interpolated new landmarks to avoid irregular deformation. In the existing work of point-driven facial animation, RBF network [24], [25] is trained to map the source points to the template mesh so as to maintain the topology of template. However, in our case, it is beneficial to keep the structure of target face, hence, we only perform rigid alignment instead.

For mesh deformation, we employ bounded biharmonic blending [26], in which the blending weights are obtained when minimizing the Laplacian energy subject to bound constraints.

B. Rigid Alignment

To provide a good initialization for subsequent fittings, we perform a multi-stage alignment as in [27]. This procedure aligns the facial data with the annotated model, during which a rigid transform would be estimated. There are three steps in the multi-stage alignment [27], with each step taking the result of previous step as input.
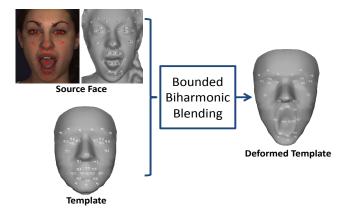
Fig. 2: Procedure of feature point-driven mesh deformation.

Spin images [28]: The first step matches the spin images of selected raw data points to those of the model points, and finds the initial correspondence between them. After the correspondence is established, a plausible transformation would be computed and verified.

Iterative Closest Point (ICP) [29]: During ICP procedure, boundary points on the $target$ surface are disregarded, hence, residual error from the non-overlapping part of two clouds would not be counted. A trimmed ICP algorithm [27], [30] is used when ICP fails to generate a satisfied result.

Simulated annealing on Z-Buffers [31], [32]: To refine the alignment result given by previous two steps, minimization in the differences between the model and data on z-buffers is performed. Enhanced Simulated Annealing (ESA) [31], [32] is adopted for this purpose.

C. Dynamic Subdivision for Face Fitting

The next stage is dynamic fitting on the rigidly aligned face. It starts with the deformed AFM, and then subdivides the model using Loop subdivision [33]. The subdivided model is segmented into predefined face parts, on which we will perform our local fitting. We perform such divide-and-fit operation only for two levels, due to the sufficient denseness of second level. In this paper, we propose the Active Non-rigid ICP that incorporates statistical shape prior during the fitting. We will explain it in Section IV.

Loop subdivision: Different from regular parametric surface splines (e.g. B-spline, Bezier and NURBS), subdivision surface can handle shapes of arbitrary topology, and offer smoothness, flexibility and scalability [34]. Starting with a given polygonal mesh called $control\ mesh$, it subdivides the mesh following a refinement scheme. Based on the neighbouring old vertices, new vertices and faces are created. In this paper, we use Loop subdivision scheme [33] for its simplicity and efficiency. This method could work on the triangular mesh that has extraordinary vertex (valence $\neq 6$), to compute the new vertex, it only uses the vertices lying within one-neighbourhood area of this vertex. It could produce a limit surface with C^2 smoothness everywhere except at extraordinary points, where only C^1 continuity is guaranteed.

Dynamic local fitting: We follow a coarse-to-fine approach for the fitting procedure. After the initial NICP [5] fitting, for the first subdivision level, we apply Loop subdivision scheme to the fitted AFM. As for such denser model, it is neither efficient nor feasible to perform the same global fitting, hence, we propose to fit the surface locally based on the predefined face parts in AFM, which allows more local variations to be captured, resulting in a highly detailed representation of raw data. We repeat the same procedure to the first level result for the second level fitting. To alleviate the boundary discrepancies between parts, we take the mean value of the shared boundary vertices. In the next section, we will describe the method used for the local fitting.

IV. ACTIVE NONRIGID ICP

To capture more local variations, we perform local fitting based on the segmented template of subdivision levels. Inspired by the recent success in region-based face modelling [21], we employ a statistical shape model in non-rigid ICP algorithm, and propose to solve the optimal mesh controlling parameters in an alternating manner. We refer to this method as Dynamic Active Non-rigid ICP (**DA-NICP**) in this paper.

A. Local Face Part Modelling

We denote the model with N vertices $\mathcal{V}$ and M edges $\mathcal{E}$ by $\mathcal{S} = (\mathcal{V}, \mathcal{E})$, and the target surface by $\mathcal{T}$. Assume that we have P independent parts for the current face template $\mathcal{V} = \{\mathbf{v}^i\}_{i=1}^P$, the ith part $\mathbf{v}_i$ can be modelled by the following equation:

$$\mathbf{v}^i = \mathbf{B}^i \mathbf{c}^i + \mathbf{m}^i, \tag{1}$$

where $\mathbf{v}^i \in \mathfrak{R}^{4N \times 1}$ that contains 3D coordinates (x, y, z) plus homogeneous coordinates of all vertices for the corresponding part; $\mathbf{B}^i \in \mathfrak{R}^{4N \times h}$ consists of h linear bases of the PCA model, $\mathbf{m}^i \in \mathfrak{R}^{4N \times 1}$ is the mean value, and $\mathbf{c}^i \in \mathfrak{R}^{h \times 1}$ is the latent variable controlling deformation of the model. By solving for the optimal parameter $\mathbf{c}^i$, the local face region can be faithfully reconstructed and used in the subsequent fitting steps.

B. Active Non-rigid ICP Formulation

The goal of our fitting is to find a set of affine parameters $\mathcal{X} = \{\mathbf{X}^i\}_{i=1}^P$ and non-rigid parameter $\mathcal{C} = \{\mathbf{c}^i\}_{i=1}^P$ such that for each face part the displaced source vertices $\mathcal{V}(\{\mathbf{X}^i, \mathbf{c}^i\}_{i=1}^P)$ fit in the target surface. The refined correspondence, together with $\{\mathbf{X}^i, \mathbf{c}^i\}_{i=1}^P$ and the topology of model, form the parametric version of the 3D scan. Here, $\mathbf{X}^i := [\mathbf{X}_1^i \ ... \ \mathbf{X}_{n_i}^i]^T$ consists of an affine 3×4 transform matrix for each template vertex. In order to solve for the optimal affine and non-rigid parameters, we formulate two cost functions.

Optimizing affine parameters: We define the first cost function by the distance and stiffness term.

The $distance$ term measures the euclidean distance between source vertex and its closest vertex in target mesh. Given that we have the optimal non-rigid parameters and fixed correspondences between the template $\mathbf{v}^i$ and target $\mathbf{u}^i$ ($\mathbf{u}^i \in$

$\mathfrak{R}^{3N\times1}$) after the closest point matching step for each part, the distance term is defined as:

$$E_d(\{\mathbf{X}^i\}_{i=1}^P) := \sum_{\mathbf{v}^i\in\mathcal{V}}\sum_{j=1}^{n_i}\|\mathbf{X}_j^i\mathbf{v}_j^i-\mathbf{u}_j^i\|^2$$

$$= \sum_{\mathbf{c}^i\in\mathcal{C}}\sum_{j=1}^{n_i}\|\mathbf{X}_j^i(\mathbf{B}_j^i\mathbf{c}_j^i+\mathbf{m}_j^i)-\mathbf{u}_j^i\|^2$$

$$= \sum_{i=1}^P\left\|\begin{bmatrix}\mathbf{X}_1^i & & \\ & \ddots & \\ & & \mathbf{X}_{n_i}^i\end{bmatrix}\begin{bmatrix}\widetilde{\mathbf{v}}_1^i\\ \vdots\\ \widetilde{\mathbf{v}}_{n_i}^i\end{bmatrix}-\begin{bmatrix}\mathbf{u}_1^i\\ \vdots\\ \mathbf{u}_{n_i}^i\end{bmatrix}\right\|^2, \quad (2)$$

where $\widetilde{\mathbf{v}}_j^i = \mathbf{B}_j^i\mathbf{c}_j^i + \mathbf{m}_j^i$. Note that in [5], due to potential missing data in the scan, a reliability weight is defined for each correspondence pair to improve the robustness. Since the use of linear shape model helps to regularize the fitting, we choose not to include this additional weight here.

The current formulation is not easy to differentiate, thus we need to swap the position of unknown term $\mathbf{X}$ and current vertices $\mathbf{v}$ to obtain a standard form. Let us define $\mathbf{U}^i := [\mathbf{u}_1^i,...,\mathbf{u}_{n_i}^i]^T$, and create sparse matrix $\mathbf{D}^i$ with diagonal blocks being the transpose of each vertex $\widetilde{\mathbf{v}}_j^i$, Equation 2 can then be rearranged as:

$$E_d(\{\mathbf{X}^i\}_{i=1}^P) = \sum_{i=1}^P\left\|\mathbf{D}^i\mathbf{X}^i-\mathbf{U}^i\right\|_F^2. \quad (3)$$

The $stiffness$ term penalises differences between the transformation matrices assigned to neighbouring vertices, therefore, enforces a constraint on neighbouring vertices. This term can defined as [5]:

$$E_s(\{\mathbf{X}^i\}_{i=1}^P) = \sum_{i=1}^P\left\|(\mathbf{M}^i\otimes\mathbf{G}^i)\mathbf{X}^i\right\|_F^2. \quad (4)$$

For the ith face part, $\mathbf{G}^i := (1,1,1,\gamma^i)$, γ^i is used to balance the scale of rotational and skew factor against the translational factor. It depends on the units of the data and the deformation type to be expressed. $\mathbf{M}^i$ is the node-arc incidence matrix of the template mesh topology [5].

After defining these two terms, the complete cost function is a quadratic function:

$$E_X(\{\mathbf{X}^i\}_{i=1}^P) := E_d + \alpha E_s$$

$$= \sum_{i=1}^P\left\|\begin{bmatrix}\alpha\mathbf{M}^i\otimes\mathbf{G}^i\\ \mathbf{D}^i\end{bmatrix}\mathbf{X}^i-\begin{bmatrix}\mathbf{0}\\ \mathbf{U}^i\end{bmatrix}\right\|_F^2$$

$$= \sum_{i=1}^P\left\|\mathbf{A}^i\mathbf{X}^i-\widetilde{\mathbf{U}}^i\right\|_F^2, \quad (5)$$

which is a well-known linear least square problem. The minimum occurs where the gradient vanishes, that is $[\partial E_X/\partial\mathbf{X}^1;...;\partial E_X/\partial\mathbf{X}^P] = \mathbf{0}$. Thus Equation 5 has closed-form solution for each part:

$$\mathbf{X}^i = (\mathbf{A}^{iT}\mathbf{A}^i)^{-1}\mathbf{A}^{iT}\widetilde{\mathbf{U}}^i. \quad (6)$$

Here, α is the stiffness weight that influences the flexibility of the template. Therefore, for each iteration, given a set of

non-rigid parameters, we determine the optimal deformation in the sense that it exactly minimises the cost function for fixed stiffness and correspondences.

Optimizing non-rigid parameters: Assume that we already have the optimal affine parameters $\{\mathbf{X}^i\}_{i=1}^P$, we need to find the best $\{\mathbf{c}^i\}_{i=1}^P$. Based on the euclidean distance between closest points and template vertices, this problem can be formulated as:

$$E_c(\{\mathbf{c}^i\}_{i=1}^P) := \sum_{\mathbf{X}^i\in\mathcal{X}}\sum_{j=1}^{n_i}\|\mathbf{X}_j^i\mathbf{v}_j^i-\mathbf{u}_j^i\|^2 \quad (7)$$

$$= \sum_{\mathbf{X}^i\in\mathcal{X}}\sum_{j=1}^{n_i}\|\mathbf{X}_j^i(\mathbf{B}_j^i\mathbf{c}_j^i+\mathbf{m}_j^i)-\mathbf{u}_j^i\|^2$$

$$= \sum_{i=1}^P\left\|\widetilde{\mathbf{X}}^i(\mathbf{B}^i\mathbf{c}^i+\mathbf{m}^i)-\mathbf{u}^i\right\|^2,$$

and

$$\widetilde{\mathbf{X}}^i = \begin{bmatrix}\mathbf{X}_1^i & & \\ & \ddots & \\ & & \mathbf{X}_{n_i}^i\end{bmatrix}.$$

To solve this equation, we take the partial derivative with regard to each $\mathbf{c}^i$ and take the minimum when it approaches to zero:

$$(\widetilde{\mathbf{X}}^i\mathbf{B}^i)^T(\widetilde{\mathbf{X}}^i\mathbf{B}^i)\mathbf{c}^i+(\widetilde{\mathbf{X}}^i\mathbf{B}^i)^T(\mathbf{X}^i\mathbf{m}^i-\mathbf{u}^i)=\mathbf{0},$$

yielding the closed-form solution,

$$\mathbf{c}^i = -[(\widetilde{\mathbf{X}}^i\mathbf{B}^i)^T(\widetilde{\mathbf{X}}^i\mathbf{B}^i)]^{-1}(\widetilde{\mathbf{X}}^i\mathbf{B}^i)^T(\mathbf{X}^i\mathbf{m}^i-\mathbf{u}^i). \quad (8)$$

C. Alternating Optimization Approach

Given Equation 5 and 7, our goal is to find the best combination of affine and non-rigid parameters that aligns the template towards the raw scan. However, it is not feasible to solve both parameters simultaneously in each iteration. To account this problem, we propose to solve two different equations in an alternating manner, one for $\mathbf{X}$ and one for $\mathbf{c}$, as:

$$\mathbf{X}_o = \operatorname*{argmin}_{\mathbf{X}}\sum_{i=1}^P\left\|\mathbf{A}_c^i\mathbf{X}^i-\widetilde{\mathbf{U}}_c^i\right\|_F^2, \quad (9)$$

$$\mathbf{c}_o = \operatorname*{argmin}_{\mathbf{c}}\sum_{i=1}^P\left\|\widetilde{\mathbf{X}}_o^i(\mathbf{B}^i\mathbf{c}_c^i+\mathbf{m}^i)-\mathbf{u}_c^i\right\|^2. \quad (10)$$

Since the optimal solution for each part does not depend on the others (Equations 6 and 8), we perform the individual fitting on each of them. Given an initial estimate of non-rigid parameter $\mathbf{c}_c^i$, we can form $\mathbf{A}_c^i$ and solve optimal $\mathbf{X}_o^i$ using Equation 6. Given the optimum $\mathbf{X}_o^i$, we can construct $\widetilde{\mathbf{X}}_o^i$ and find the optimal $\mathbf{c}_o^i$ using Equation 8. As such, for this iteration, the optimal $\mathbf{X}_o^i$ and $\mathbf{c}_o^i$ are considered to be found. The alternating optimization for each face part is described in Algorithm 7.

V. EXPERIMENTS

Apart from visual comparison between fittings, we conduct two experiments to provide quantitative measures of

Algorithm 1: Alternating optimization on Active Non-rigid ICP

Require: Two point clouds, i^{th} part $\mathbf{v}^i$ of face template $\mathcal{V}$ and target $\mathcal{T}$.

1 Initialize the non-rigid parameters $\mathbf{c}_0^i$ and affine parameters $\mathbf{X}_0^i$.
2 **for** *each stiffness* $\alpha_i \in \{\alpha_1, ..., \alpha_n\}$, $\alpha_i > \alpha_{i+1}$ **do**
3   **while** $\|\mathbf{X}_j^i - \mathbf{X}_{j-1}^i\| \geq \varepsilon$ **do**
4   Find preliminary correspondences for $\mathcal{V}(\mathbf{X}_{j-1}^i, \mathbf{c}_{j-1}^i)$.
5   Get optimal affine transform $\mathbf{X}_j^i$ (Eqn. 6) from $\mathbf{c}_{j-1}^i$, preliminary correspondences and α_i.
6   Get optimal non-rigid parameters $\mathbf{c}_j^i$ (Eqn. 8) from $\mathbf{X}_j^i$.

Return $\mathbf{X}^i$

our proposed dynamic Active Non-rigid ICP (**DA-NICP**) algorithm. To demonstrate the advantage of putting subspace constraint on fitting, we introduce **D-NICP** - a method similar to DA-NICP that uses dynamic subdivision surfaces and perform local fitting, but with NICP [5] chosen as the only fitting strategy. The third method to compare is the deformable fitting with subdivision on AFM in [3] (we refer as **FEM**). To further evaluate the impact of point-driven mesh deformation on fitting, we perform the DA-NICP and D-NICP fitting with deformed AFM (**-PD**) and with original AFM.

The first experiment shows that the proposed point-driven mesh deformation helps to prevent incorrect fitting of mouth, thus D-NICP-PD and DA-NICP-PD are able to cover the majority of mouth region. The second experiment is a facial expression recognition (FER) experiment on BU-4DFE database [35]. This experiment proves that an accurate fitting can benefit the recognition task (accurate dense alignment is important for accurate face and facial expression recognition [36], [37], [38], [27]), even when a simple feature extraction and recognition method is used.

A. Experiment Setup

Linear shape models: To train our PCA model for each face part, we start with the point-driven deformed AFM, and run D-NICP-PD fitting on the BU-3DFE database [39]. We align the instances of individual part and train a shape model on them, keeping 95% of the variations. In total, there are two levels of shape model - 6 models for the first subdivision level, and 13 for the second level.

FER method: Based on the annotated parts, we extract the main face region that contains eyes, mouth, cheeks and nose from the converted depth image. In order to further demonstrate the advantage of using the deformed AFM, we also conducted FER experiments using only mouth region. We divided the extracted region into non-overlapping blocks, for which HONV (Histogram of Oriented Normal Vectors) [40] features are computed (to this end other simple histogram-based features can be used [41]). The final feature descriptor is the concatenation of feature vectors from the corresponding blocks.

Since the goal of our experiment is to compare the different fitting methods, it is less important to select between advanced recognition methods. We choose the multi-class SVM classifier with one-against-one approach [42]. Radial Basis Function (RBF) kernel is employed as the distance metric. An empirical grid search is performed over the parameters in RBF kernel.

FER experiment protocol: We create a ten-fold partition of the data, each time one fold is used for testing, all the others are used for training. Our BU-4DFE experiment is sequence-independent, which means the query sequence never appears in the training set.
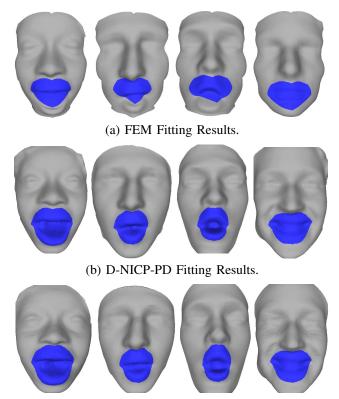
B. Mouth Coverage Experiment

To evaluate the mouth coverage of proposed fitting procedures, we select 100 scans that have exaggerated expressions of different subjects in BU-4DFE [35] and we manually annotated the mouth region. We extract the manually annotated (G) and detected (D) mouth regions specified by the alpha hull and measure tracking accuracy as $S = 1 - \frac{A(D \cap G)}{A(D \cup G)}$. $A(\cdot)$ describes the 2D area of a region. Note that the smaller S the more overlap we have. Table I shows the percentages of covered area for different fitting methods. As can be observed, both D-NICP-PD (0.2891) and DA-NICP-PD (0.3058) better cover the mouth region than those without using deformed templates. The result suggests that the mesh deformation driven by 2D face alignment greatly helps subsequent 3D fitting in locating correct face part. And this procedure can be highly efficient, hence, the time taken in this stage can be neglected. See Fig. 3 for qualitative examples of annotated mouth region.

Method	Mean S value
FEM	0.4086
D-NICP-PD	**0.2891**
D-NICP	0.3841
DA-NICP-PD	0.3058
DA-NICP	0.3688

TABLE I: Mean S values of 100 annotated scans.

C. BU-4DFE Experiment

BU-4DFE database consists of 4D faces (sequences of 3D faces), with temporal and spatial resolution being 25 frames/second and 35,000 vertices, respectively. It includes 101 subjects each containing sequences of the six proto-typical facial expressions. FER experiment on BU-4DFE is challenging, since more than one instance per subject are labelled as a specific expression, while those instances differ slightly from each other. For our experiment, we pick the apex frames for each sequence from temporal segment annotation provided in [37]. Due to the fact that each sequence has different apex period, the numbers of instances per expression differ from each other. An unbalanced training set would affect the quality of classifier, therefore, we trim

(a) FEM Fitting Results.

(b) D-NICP-PD Fitting Results.

(c) DA-NICP-PD Fitting Results.

Fig. 3: Sample fitting results with annotated mouth region highlighted (in blue).
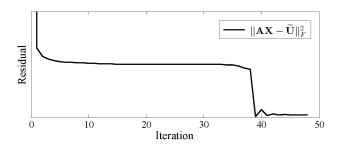

Fig. 4: Example residual error changes as the dynamic Active Non-rigid ICP (DA-NICP) fitting of the upper lip part in the second subdivision level progresses.

As can be seen, the residual error monotonically decreases and gradually converges to a minimum value.

D. Visualization of Fitting Results

We show sample BU-4DFE fitting results of previously mentioned methods, one example for each expression (see Fig. 5). We manually crop some images for a better demonstration. It is obvious to see that the proposed methods model the face better than FEM, especially for mouth region, where FEM produces distorted and weird shapes. For expression surprise (row 6), FEM fails to capture the chin and full lower lip. Since our fitting procedures are dynamic, the template would be stretched to cover the whole face, hence, we can fully represent the original face structure. The main difference between fittings with and without deformed AFM is the position of annotated parts (we demonstrated it in previous experiment). In most cases, without the deformed AFM, D-NICP and DA-NICP can still fit the entire face, therefore, it is difficult to spot great difference from the ones with deformed template in Fig. 5.

Comparing to FEM, the DA-NICP-PD and D-NICP-PD capture more details including wrinkles and eyelids. This is because D-NICP-PD has less constraint on the shape deformation, hence, allowing very closely fitting to the scan surface. While DA-NICP-PD regularizes the shape using PCA model, as a result, produces a relatively smooth fitting. One major problem of D-NICP-PD is that the boundaries between each face part are not very consistent, since the constraint on the deformation of border vertices is weak. An typical example can be found in surprise example (row

the data for each apex sequence and generate a balanced set that consists of roughly 6000 instances, with each expression having nearly 1000 instances.

Table II shows the performance measures of the five fitting methods. As can be observed, DA-NICP-PD consistently outperforms the others in both main face and mouth experiments, achieving overall accuracies of 81.16% and 80.41% respectively. Not surprisingly, owing to a better annotated mouth region, methods that utilize deformed AFM (D-NICP-PD and DA-NICP-PD) gain more accuracy than those with original template, and the margin can be 2.7% and even higher. Although we expect a decreased performance when using only mouth for FER, other than the huge decrease of FEM (from 75.87% to 73.38%), all our methods drop less in performance. In particular, DA-NICP-PD only drop 0.75% in accuracy.

Furthermore, it is important to notice that even without deformed AFM, our fitting methods (D-NICP and DA-NICP) show improvement of 0.87% and 2.56% correspondingly. It proves that a more accurate fitting would benefit the FER, and the proposed fitting methods manage to capture facial motion that is informative for FER. However, an over detailed fitting as D-NICP will be sensitive to the noise in scan, thus by reducing the motion caused by the noise, DA-NICP achieves a better performance. In addition, DA-NICP has very nice convergence properties. In Fig. 4, we show one example of residual error changes as the fitting progresses.

Method	Main face		Mouth	
	F_1[%]	CR[%]	F_1[%]	CR[%]
FEM	75.65	75.87	73.49	73.38
D-NICP-PD	79.24	79.47	78.18	78.13
D-NICP	76.50	76.74	74.74	75.05
DA-NICP-PD	**80.94**	**81.16**	**80.44**	**80.41**
DA-NICP	78.22	78.43	76.15	76.57

TABLE II: Classification rate (CR), and average F1-measure (F_1) achieved with all the methods and face regions (i.e. main face, mouth) in BU-4DFE (%). The best performance achieved for each measure is show in bold.

6, column 5) of D-NICP-PD, where there are black holes around the mouth, this suggests inconsistency between mouth and neighbouring parts.

VI. CONCLUSION

We propose a dynamic local fitting procedure that gains benefits from dynamic subdivision framework, and show how to adopt the NICP algorithm to our procedure. The proposed fitting procedure is shown capable of modelling subtle facial details. More importantly, we present a statistical model for describing the faces and we combined it with NICP for 3D face alignment. We have shown that the proposed algorithm largely outperforms state-of-the-art 3D facial deformable models, such as the ones that use Finite Element Methods (FEM), especially when it comes to the description and alignment of the mouth region. Furthermore, to avoid incorrect fitting of face part (especially mouth), we propose a pre-processing procedure where the feature points, taken from a robust 2D face alignment procedure, are employed to deform template. Extensive experimental results verify our theoretical expositions.

ACKNOWLEDGEMENT

REFERENCES

[1] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *ACM SIGGRAPH*, 1999, pp. 187–194.

[2] P. Paysan, R. Knothe, S. Romdhani B. Amberg, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," in *AVSS*, Genova, Italy, 2009.

[3] G. Passalis, I.A. Kakadiaris, and T. Theoharis, "Intraclass retrieval of nonrigid 3d objects: Application to face recognition," *T-PAMI*, vol. 29, no. 2, pp. 218–229, 2007.

[4] Y. Wang, G. Pan, and Z. Wu, "3d face recognition in the presence of expression: A guidance-based constraint deformation approach," in *CVPR*, June 2007, pp. 1–7.

[5] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid icp algorithms for surface registration," in *CVPR*, June 2007, pp. 1–8.

[6] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: Reconstruction and parameterization from range scans," *TOG*, vol. 22, no. 3, pp. 587–594, July 2003.

[7] G. Pan, X. Zhang, Y. Wang, Z. Hu, X. Zheng, and Z. Wu, "Establishing point correspondence of 3d faces via sparse facial deformable model," *TIP*, vol. 22, no. 11, pp. 4170–4181, Nov 2013.

[8] D.C. Schneider and P. Eisert, "Fast nonrigid mesh registration with a data-driven deformation prior," in *ICCV Workshops*, Sept 2009, pp. 304–311.

[9] I.A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal face recognition: combination of geometry with physiological information," in *CVPR*, June 2005, pp. 1022–1029 vol. 2.

[10] B. Amberg, R. Knothe, and T. Vetter, "Expression invariant 3d face recognition with a morphable model," in *FG*, Sept 2008, pp. 1–6.

[11] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *CVPR*, Portland, Oregon, USA, June 2013.

[12] H. Li, R. W. Sumner, and M. Pauly, "Global correspondence optimization for non-rigid registration of depth scans," *SGP*, vol. 27, no. 5, July 2008.

[13] Z. Wang, M. Grochulla, T. Thormahlen, and H.-P. Seidel, "3d face template registration using normal maps," in *3DV*, June 2013, pp. 295–302.

[14] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3d face recognition in the presence of facial expressions: an annotated deformable model approach," in *CVPR Workshops*, June 2005, pp. 171–171.

[15] C. Mandal, H. Qin, and B. C. Vemuri, "A novel fem-based dynamic framework for subdivision surfaces," in *SM*, New York, NY, USA, 1999, SMA '99, pp. 191–202, ACM.

[16] G. Passalis, P. Perakis, T. Theoharis, and I.A. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3d face recognition," *T-PAMI*, vol. 33, no. 10, pp. 1938–1951, 2011.

[17] T. Fang, X. Zhao, O. Ocegueda, S. K Shah, and I.A. Kakadiaris, "3d/4d facial expression analysis: an advanced annotated face model approach," *IVC*, vol. 30, no. 10, pp. 738–749, 2012.

[18] D. Hahnel, S. Thrun, and W. Burgard, "An extension of the icp algorithm for modeling nonrigid objects with mobile robots," in *IJCAI*, San Francisco, CA, USA, 2003, pp. 915–920, Morgan Kaufmann Publishers Inc.

[19] T. Albrecht, M. Luthi, and T. Vetter, "A statistical deformation prior for non-rigid image and shape registration," in *CVPR*, June 2008, pp. 1–8.

[20] C. David Schneider and Peter Eisert, "Algorithms for automatic and robust registration of 3d head scans," *JVRB*, vol. 7(2010), no. 7, 2010.

[21] J. R. Tena, F. De la Torre, and I. Matthews, "Interactive region-based linear 3d face models," in *ACM SIGGRAPH*, New York, NY, USA, 2011, pp. 76:1–76:10.

[22] A. Asthana, S. Zafeiriou, G. Tzimiropoulos, S. Cheng, and M. Pantic, "From pixels to response maps: Discriminative image filtering for face alignment in the wild," *T-PAMI*, 2015.

[23] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *CVPR*, 2014.

[24] L. Dutreve, A. Meyer, and S. Bouakaz, "Feature points based facial animation retargeting," in *VRST*, New York, NY, USA, 2008, pp. 197–200, ACM.

[25] T. C. S. Rendall and C. B. Allen, "Reduced surface point selection options for efficient mesh deformation using radial basis functions," *J. Comput. Phys.*, vol. 229, no. 8, pp. 2810–2820, Apr. 2010.

[26] A. Jacobson, I. Baran, J. Popović, and O. Sorkine, "Bounded biharmonic weights for real-time deformation," in *ACM SIGGRAPH*, New York, NY, USA, 2011, pp. 78:1–78:8, ACM.

[27] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *T-PAMI*, vol. 29, no. 4, pp. 640–649, April 2007.

[28] A. Johnson, *Spin-Images: A Representation for 3-D Surface Matching*, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[29] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *T-PAMI*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[30] D. Chetverikov, D. Stepanov, and P. Krsek, "Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm," *IVC*, vol. 23, pp. 299–309, 2005.

[31] P. Siarry, G. Berthiau, F. Durdin, and J. Haussy, "Enhanced simulated annealing for globally minimizing functions of many-continuous variables," *TOMS*, vol. 23, no. 2, pp. 209–228, June 1997.

[32] G. Papaioannou, E. A. Karabassi, and T. Theoharis, "Reconstruction of three-dimensional objects through matching of their parts," *T-PAMI*, vol. 24, no. 1, pp. 114–124, Jan 2002.

[33] C. Loop, "Smooth Subdivision Surfaces Based on Triangles," M.S. thesis, The University of Utah, Aug. 1987.

[34] J. Warren and H. Weimer, *Subdivision Methods for Geometric Design: A Constructive Approach*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2001.

[35] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database," in *FG*, Sept 2008, pp. 1–6.

[36] G. Rajamanoharan, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3d facial expression recognition: A comprehensive survey," *IVC*, vol. 30, no. 10, pp. 683–697, 2012.

[37] G. Rajamanoharan, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3d facial expression dynamics," *IVC*, vol. 30, no. 10, pp. 762–773, 2012.

Fig. 5: Qualitative fitting results from BU-4DFE database. Each row shows examples from one of the six prototypical expressions (in order of $angry$, $disgust$, $fear$, $happy$, sad and $surprise$). Column 1 shows the raw face scan, the second to sixth columns illustrate the shapes from the FEM, and the proposed approaches D-NICP, DA-NICP, D-NICP-PD and DA-NICP-PD, respectively.

[38] I. Marras, S. Zafeiriou, and G. Tzimiropoulos, "Robust learning from normals for 3d face recognition," in *ECCV Workshops*, 2012, pp. 230–239.

[39] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato, "A 3d facial expression database for facial behavior research," in *FG*, April 2006, pp. 211–216.

[40] S. Tang, X. Wang, X. Lv, T. X. Han, J. Keller, Z. He, M. Skubic, and S. Lao, "Histogram of oriented normal vectors for object recognition with a depth sensor," in *ACCV*, Berlin, Heidelberg, 2013, pp. 525–538, Springer-Verlag.

[41] G. Rajamanoharan, S. Zafeiriou, and M. Pantic, "Binary pattern analysis for 3d facial action unit detection," in *BMVC*, Guildford, UK, September 2012.

[42] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM TIST*, vol. 2, no. 3, pp. 27:1–27:27, 2011.