



Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Statistical non-rigid ICP algorithm and its application to 3D face alignment[☆]

Shiyang Cheng^{a,*}, Ioannis Marras^a, Stefanos Zafeiriou^a, Maja Pantic^{a,b}

^aDepartment of Computing, Imperial College London, United Kingdom

^bEEMCS, University of Twente, Netherlands

ARTICLE INFO

Article history:

Received 3 October 2015

Received in revised form 13 October 2016

Accepted 14 October 2016

Available online xxx

Keywords:

3D face alignment

Statistical shape model

Non-rigid ICP algorithm

BU-4DFE database

FRGC v2 database

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 it 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.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Three-dimensional representation of face has always been a valuable source for face recognition and facial behaviour analysis. The huge descriptive power and rich shape variation of 3D human face make it much more informative than the corresponding 2D projection. Due to its complex nature, the structure and topology of the 3D facial scan varies from frame to frame. To obtain a relatively consistent facial representation, it is always preferable to use a 3D deformable facial model. Thus, the building and fitting of the 3D deformable model becomes a key step in retrieving 3D facial information, and not surprisingly, it has received a lot of attention in the past decade [1–8]. Owing to the recent development on cost-effective depth cameras, such as Kinect 2, Creative Senz3D™ and Intel® RealSense™ Camera (F200), further attention has been drawn on this field of study.

3D face fitting, also referred to as 3D face registration [5,8,9] in the literature, aims at aligning two sets of point clouds or meshes,

and overlapping the face template as close as possible with the target surface (i.e. facial scan). In the scope of 3D deformable model fitting, the face template refers to the 3D deformable face model. 3D Morphable Model (3DMM) [1] is the most commonly used technique to introduce prior knowledge on 3D human face. 3DMM has been widely adopted in various 3D fitting methods [2,10–12], since it can well estimate an unseen 3D facial shape by solving the shape, texture, pose and illumination parameters simultaneously. Fitting of 3D facial models is very important, since they can be used to identify particular facial landmarks or to recognize and define facial deformations that benefits face recognition [3,13] and facial performance transfer [14]. Furthermore, accurate fitting and modelling of human expressions is able to boost the performance of facial expression recognition [15–17].

Despite the importance, the majority of existing methodologies use just a simple pre-defined mesh model, parameterised or not, to fit a target mesh [2,3,5,8,11,18]. In [2,11], in order to handle recognition under different pose and lighting conditions, a 3DMM that separates parameters for shape, head pose and illumination is employed. Inspired by this, Amberg et al. [10] developed an expression-invariant 3D face recognition algorithm, in that, they fit an identity/expression separated 3DMM to the facial scan and normalize the resulting face by removing the pose and expression components [10]. Unfortunately, these methodologies may fail to describe properly the complex, non-linear and highly deformable

[☆] This paper has been recommended for acceptance by Sinisa Todorovic.

* Corresponding author.

E-mail address: shiyang.cheng11@imperial.ac.uk (S. Cheng).

URL: <http://ibug.doc.ic.ac.uk/people/scheng> (M. Pantic).

structure of the face. Especially when fitting the data captured by high-resolution face capturing system, such as Di3D (Dimensional Imaging [19]), the results are likely to be over-smooth and thus lose important facial details and micro-expression.

In this paper, we examine the problem of fitting a 3D facial models to high-resolution depth scan. Our key contribution is a new active method for describing and fitting 3D faces, which is achieved by learning a set of local statistical model for facial parts, and combining them with the non-rigid Iterative Closest Point (ICP) algorithm [5]. Besides it, we propose a dynamic local fitting procedure that makes full use of dynamic subdivision framework. To this end, we successfully adopt the proposed active method in the fitting procedure, and show that it manages to accurately model the subtle facial feature. Additionally, we provide a point-driven mesh deformation procedure in the data pre-processing stage that helps to prevent incorrect facial part fitting. It deforms 3D template model under the guidance of the state-of-the-art 2D face alignment algorithm [20].

The remainder of the paper is organized as follows. Section 2 gives a brief introduction to the existing works that are related to this paper, and a short discussion over the merits and demerits of them. Section 3 explains our dynamic subdivision framework as well as the local fitting procedure. Next, in Section 4, we describe the proposed novel Active Non-rigid ICP algorithm. In Section 5, we show the results of our experiments followed by an in-depth discussion. Lastly, we conclude the paper in Section 6.

2. 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 [5,21,22] and normal maps [23]. In this section, we will give a brief introduction to different kinds of 3D fitting algorithms.

2.1. Annotated deformable model fitting

Passalis et al. [3] proposed a deformable fitting framework based on AFM (Annotated Face Model) [18,24], which has been one of the state-of-the-art 3D fitting methods for face recognition. This method defines mass matrix \mathbf{M} , damping matrix \mathbf{D} and stiffness matrix \mathbf{K} to control the deformation of template. In this physics-based model, the mass and damping matrix relate to the kinetic energy and energy dissipation, while the stiffness matrix \mathbf{K} represents the strain energy, which determines the elastic properties. Assume that \mathbf{q} denotes the n degrees of freedom in the template model, its analytical formulation can be written as:

$$\mathbf{M} \frac{d^2 \mathbf{q}}{dt} + \mathbf{D} \frac{d\mathbf{q}}{dt} + \mathbf{K}\mathbf{q} = \mathbf{f}_q,$$

and \mathbf{f}_q describes the external force that is proportional to the Euclidean distance between each vertex in limit surface and its closest point in target mesh. 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 [25] 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 [26]. 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 [27]. In [27] in order to better capture facial expressions an extra fitting strategy using Thin Plate Splines (TPS) was applied.

2.2. Non-rigid iterative closest point

Another family of 3D fitting algorithms that uses a single template is non-rigid ICP (Iterative Closest Point), where correspondence of points is 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. [28] proposed a hierarchical method that firstly fits a skeleton structure to the scan points and then aligns local parts. Allen et al. [6] assumed 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.

2.3. Statistical shape models fitting

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 [7,8,10,29–31]. Albrecht et al. [29] proposed to use Tikhonov regularization that is based on a statistical deformation model to regularize the shape. Bolkart and Wuhrer [32] proposed to separately model the motion of 3D facial shape and its identity in a multilinear statistical model. 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. Schneider et al. [8] further extended to the problem of head scan registration [30]. Similar to [8,30], 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, to assist our methods modelling facial expression, we provide a point-driven mesh deformation procedure that generates deformed reference shape based on robust 2D face alignment.

3. Dynamic subdivision framework

The core idea of our framework is to dynamically fit facial data using a deformable 3D face model, and to provide an accurately fitted surface. In contrast to previous works on 3D surface registration [3,18,24] 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 fitting accuracy, thus loses essential details (for example, wrinkles in the forehead) for the subsequent expression recognition task. Our framework mainly consists of three processing stages: (1) Point-driven template deformation; (2) Rigid alignment; (3) Local deformable model fitting.

3.1. Point-driven template deformation

Our fitting starts from annotated model of the face (AFM) [18,24], which is segmented into several parts according to the facial physiology information [18,24]. Each part groups the vertices that have the most common characteristics and movement, therefore, lowers the difficulties of local surface registration. 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 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.

To tackle this problem, in [33], based on the locations of the known landmarks in a database, they predict the landmarks on the unseen scan, and fit the expression using a blendshape model under the guidance of those predicted landmarks. Inspired by it, we propose to deform AFM based on the fitting of 2D face alignment [20,34,35], it is reliable to deform the mesh under the guidance of 2D fitting. Therefore, we use publicly available Chehra face tracker [35], 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 [36,37] 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 [38], in which the blending weights are obtained when minimizing the Laplacian energy subject to bound constraints.

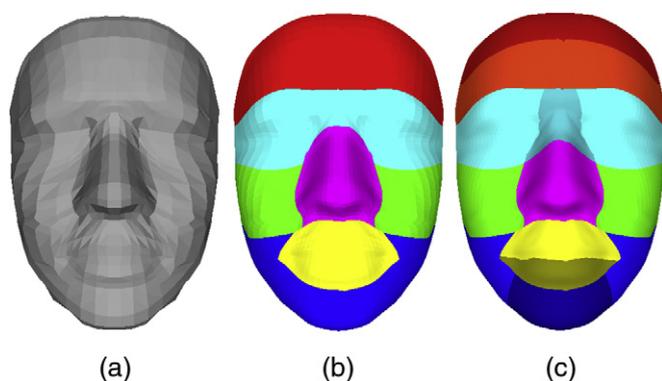


Fig. 1. Annotated areas for different levels of AFM [18,24]. (a) The starting level, without subdivision and segmentation; (b) First subdivision level, with 6 annotated face parts (cheeks are combined into one part); (c) Second subdivision level, with 13 individual parts.

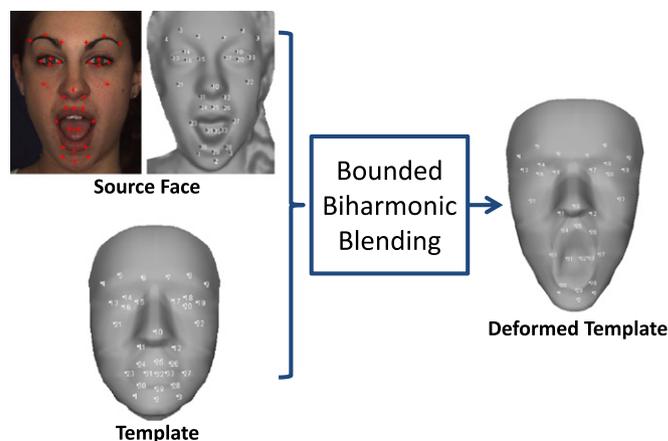


Fig. 2. Procedure of feature point-driven mesh deformation.

3.2. Rigid alignment

To provide a good initialization for subsequent fittings, we perform a multi-stage alignment as in [13]. 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 [13], with each step taking the result of previous step as input.

3.2.1. Spin Images

Spin Images [39] encodes the global properties of any surface in an object-oriented coordinate system where the description of a surface is view-independent. This algorithm is useful for correcting rotation under the premise that arbitrary rotations and translations in the database do not exist. Hence, our first step is to match 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.

3.2.2. Iterative Closest Point (ICP)

ICP [40] is the main step in our rigid alignment procedure. Given two fixed point clouds the *source* and *target*, ICP iteratively revises the transformation (combination of translation and rotation) to minimize the distance between them. Here, the 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 [13,41] is used when ICP fails to generate a satisfied result.

3.2.3. Simulated annealing on z-buffers

To refine the alignment result given by previous two steps, minimization in the differences between the model and data on z-buffers is performed. Z-buffer is essentially resampling the data without any use of triangular mesh information, this optimization could achieve a very high accuracy. Instead of using an exhaustive search, Enhanced Simulated Annealing (ESA) [42,43] is adopted to find the global minimum. Simulate Annealing (SA) is a probabilistic technique for estimating the global optimum of a cost function. The procedure of lowering temperature in SA process can be interpreted as lowering the probability of reaching a worse solution. Comparing with the original SA method, ESA introduces an adaptive cooling scheme to adjust the decreasing rate of probability based on the previous solution.

3.3. Dynamic subdivision for face fitting

The next stage is dynamic fitting on the rigidly aligned face. It starts with the deformed AFM (i.e. the model shown in Fig. 1 (a)), with which we perform a NICP [5] fitting to roughly match the target surface. After this, we subdivide the fitted model using Loop subdivision [44]. Since the subdivided meshes always have identical number of vertex and topology as the corresponding AFM (shown in Fig. 1 (b) and (c)), it is straightforward to identify each facial part by our pre-defined segmentation and perform a local fitting for each individual part. Thanks to the sufficient denseness of second level, we only need to perform such divide-and-fit operation for two levels. In this paper, we propose the Active Non-rigid ICP that incorporates statistical shape prior during the fitting, which we will explain in Section 4.

3.3.1. 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 [45]. Starting with a given polygonal mesh called *controlmesh*, it subdivides the mesh following a refinement scheme. Based on the neighbouring old vertices, new vertices and faces are created. Existing subdivision schemes can be grouped into two categories: approximating and interpolating subdivision techniques. In general, approximating subdivision offers greater smoothness. Typical approximating subdivision are Catmull-Clark [46] and Loop [44] schemes.

In this paper, we use Loop subdivision scheme [44] 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.

3.3.2. Dynamic local fitting

We follow a coarse-to-fine approach for the fitting procedure. After the initial NICP [5] fitting, we apply subdivide the fitted holistic AFM (as shown in Fig. 1 (a)) to obtain the initial template model for the first subdivision level. Considering the large number of vertices in the subdivided model, it is neither efficient nor feasible to perform the same global fitting. Therefore, we propose to fit such model locally based on the predefined face parts (see Fig. 1 (b) and (c)). In such, more local variations are likely to be captured, and thus leads to 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.

4. Statistical Non-rigid ICP algorithm

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 [31], we employ a statistical shape model in non-rigid ICP algorithm (see Section 5 for details of shape model building), 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.

4.1. Local face part modelling

We denote the model with N vertices \mathcal{V} and M edges \mathcal{E} by $S = (\mathcal{V}, \mathcal{E})$, 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 i^{th} 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.

In this paper, there will be two groups of local face part models. The first group consists of 6 local part models which model the first subdivision level, and the second ones consist of 13 part models for the second subdivision level.

4.2. 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, $\mathbf{X}^i := [\mathbf{X}_1^i \dots \mathbf{X}_{n_i}^i]^T$ consists of an affine 3×4 transform matrix for each template vertex.

Our energy function consists of a distance and a stiffness term. The *distance* term measures the Euclidean distance between source vertex and its closest vertex in target mesh. Given that we have the non-rigid parameters and fixed correspondences between the template \mathbf{v}^i and target \mathbf{u}^i ($\mathbf{u}^i \in \mathfrak{R}^{3N \times 1}$) after the closest point matching step for each part, the distance term is defined as:

$$\begin{aligned} E_d(\{\mathbf{X}^i, \mathbf{c}^i\}_{i=1}^P) &:= \sum_{\mathbf{v}^i \in \mathcal{V}} \sum_{j=1}^{n_i} \left\| \mathbf{X}_j^i \mathbf{v}^i - \mathbf{u}_j^i \right\|^2 \\ &= \sum_{\mathbf{c}^i \in \mathcal{C}} \sum_{j=1}^{n_i} \left\| \mathbf{X}_j^i (\mathbf{B}_j^i \mathbf{c}^i + \mathbf{m}_j^i) - \mathbf{u}_j^i \right\|^2 \\ &= \sum_{i=1}^P \left\| \tilde{\mathbf{X}}^i (\mathbf{B}^i \mathbf{c}^i + \mathbf{m}^i) - \mathbf{u}^i \right\|_F^2, \end{aligned} \tag{2}$$

where

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

and $\mathbf{B}_j^i \in \mathfrak{R}^{4 \times h}$ are the corresponding principle components for the j^{th} vertex in $\mathbf{B}^i = [\mathbf{B}_1^i; \mathbf{B}_2^i; \dots; \mathbf{B}_{n_i}^i]$. 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 *stiffness* term penalises differences between the transformation matrices assigned to neighbouring vertices, therefore, enforces a constraint on neighbouring vertices. This term can be defined as [5]:

$$E_s(\{\mathbf{X}^i\}_{i=1}^P) = \sum_{i=1}^P \left\| (\mathbf{M}^i \otimes \mathbf{G}^i) \mathbf{X}^i \right\|_F^2. \tag{3}$$

For the i^{th} 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].

Combining Eqs. (2) and (3), the complete cost function can be written as:

$$\begin{aligned} \arg \min_{\mathbf{X}, \mathbf{c}} E(\{\mathbf{X}^i, \mathbf{c}^i\}_{i=1}^P) &:= \arg \min_{\mathbf{X}, \mathbf{c}} E_d + \alpha E_s \\ &= \arg \min_{\mathbf{X}, \mathbf{c}} \sum_{i=1}^P \|\tilde{\mathbf{X}}^i(\mathbf{B}^i \mathbf{c}^i + \mathbf{m}^i) - \mathbf{u}^i\|_F^2 + \alpha \sum_{i=1}^P \|(\mathbf{M}^i \otimes \mathbf{G}^i) \mathbf{X}^i\|_F^2. \end{aligned} \quad (4)$$

Here, α is the stiffness weight that influences the flexibility of the template. This optimization problem is with regards to both the affine parameters and non-rigid parameters of the statistical shape model. Unfortunately, even though Eq. (4) is convex with regards to either affine parameters or non-rigid parameters, it is not convex on the combination. Hence, to account this problem, we propose to solve the two sets of parameters in an alternating manner.

More specifically, given an initial estimate of non-rigid parameter \mathbf{c}_0^i , we solve the optimal \mathbf{X}_1^i using Eq. (8). After obtaining the optimal local affine transforms \mathbf{X}_1^i , we find the optimal non-rigid parameters \mathbf{c}_1^i using Eq. (9). As such, for this iteration, the optimal \mathbf{X}_1^i and \mathbf{c}_1^i are considered to be found. Such alternating optimization for each face part is described in Algorithm 1. We will describe the optimization step for affine parameters and non-rigid parameter separately in the following subsections.

Algorithm 1. Alternating optimization on Active Non-rigid ICP.

Require: Two point clouds, i^{th} part \mathcal{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 $\alpha_i \in \{\alpha_1, \dots, \alpha_n\}$, $\alpha_i > \alpha_{i+1}$ **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 Calculate optimal local affine transform \mathbf{X}_j^i (Eqn. 8) from \mathbf{c}_{j-1}^i , preliminary correspondences and α_i .
- 6 Calculate optimal non-rigid parameters \mathbf{c}_j^i (Eqn. 9) from \mathbf{X}_j^i .
- 7 **Return** \mathbf{X}^i

4.2.1. Local affine parameters optimization

In this part, we explain the optimization step for local affine parameters in detail. Suppose that we have an initial estimate of non-rigid parameter \mathbf{c}_0 , and the current template $\mathbf{v} = \mathbf{B}\mathbf{c}_0 + \mathbf{m}$, Eq. (2) can be rewritten as:

$$E_d(\{\mathbf{X}^i\}_{i=1}^P) = \sum_{i=1}^P \|\tilde{\mathbf{X}}^i \mathbf{v}^i - \mathbf{u}^i\|_F^2. \quad (5)$$

In order to differentiate this equation, we swap the position of unknown term \mathbf{X} and current vertices \mathbf{v} to obtain a standard form. Let us define a sparse matrix \mathbf{D}^i with diagonal blocks being the transpose of each vertex \mathbf{v}^i , Eq. (5) can then be rearranged as:

$$E_d(\{\mathbf{X}^i\}_{i=1}^P) = \sum_{i=1}^P \|\mathbf{D}^i \mathbf{X}^i - \mathbf{u}^i\|_F^2. \quad (6)$$

After substituting the distance term in Eq. (4) with the above Eq. (6), the original cost function becomes a differentiable quadratic function:

$$\begin{aligned} E_X(\{\mathbf{X}^i\}_{i=1}^P) &= \sum_{i=1}^P \|\mathbf{D}^i \mathbf{X}^i - \mathbf{u}^i\|_F^2 + \alpha \sum_{i=1}^P \|(\mathbf{M}^i \otimes \mathbf{G}^i) \mathbf{X}^i\|_F^2 \\ &= \sum_{i=1}^P \left\| \begin{bmatrix} \alpha \mathbf{M}^i \otimes \mathbf{G}^i \\ \mathbf{D}^i \end{bmatrix} \mathbf{X}^i - \begin{bmatrix} \mathbf{0} \\ \mathbf{u}^i \end{bmatrix} \right\|_F^2 \\ &= \sum_{i=1}^P \|\mathbf{A}^i \mathbf{X}^i - \tilde{\mathbf{u}}^i\|_F^2, \end{aligned} \quad (7)$$

which is a well-known linear least square problem. The minimum occurs where the gradient vanishes, that is $[\partial E_X / \partial \mathbf{X}^1; \dots; \partial E_X / \partial \mathbf{X}^P] = \mathbf{0}$. Thus Eq. (7) has closed-form solution for each part:

$$\mathbf{X}^i = (\mathbf{A}^{iT} \mathbf{A}^i)^{-1} \mathbf{A}^{iT} \tilde{\mathbf{u}}^i \quad (8)$$

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.

4.2.2. Non-rigid parameters optimization

Assume that we already have the optimal affine parameters $\{\mathbf{X}_1^i\}_{i=1}^P$, we need to find the best $\{\mathbf{c}^i\}_{i=1}^P$. To solve for the optimal non-rigid parameters, we take the partial derivative of Eq. (4) with regards to each \mathbf{c}^i and obtain the minimum when it approaches to zero:

$$(\tilde{\mathbf{X}}_1^i \mathbf{B}^i)^T (\tilde{\mathbf{X}}_1^i \mathbf{B}^i) \mathbf{c}^i + (\tilde{\mathbf{X}}_1^i \mathbf{B}^i)^T (\tilde{\mathbf{X}}_1^i \mathbf{m}^i - \mathbf{u}^i) = \mathbf{0},$$

yielding the closed-form solution,

$$\mathbf{c}^i = -\left[(\tilde{\mathbf{X}}_1^i \mathbf{B}^i)^T (\tilde{\mathbf{X}}_1^i \mathbf{B}^i) \right]^{-1} (\tilde{\mathbf{X}}_1^i \mathbf{B}^i)^T (\tilde{\mathbf{X}}_1^i \mathbf{m}^i - \mathbf{u}^i). \quad (9)$$

5. Experiments

Apart from visual comparison between fittings, we conduct three experiments to provide quantitative measures of 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 modelling facial expression, 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

Table 1
Mean S values of 100 annotated scans. The best performance achieved is shown in bold.

Method	Mean S value
FEM [3]	0.4086
D-NICP-PD [5]	0.2891
D-NICP [5]	0.3841
DA-NICP-PD	0.3058
DA-NICP	0.3688

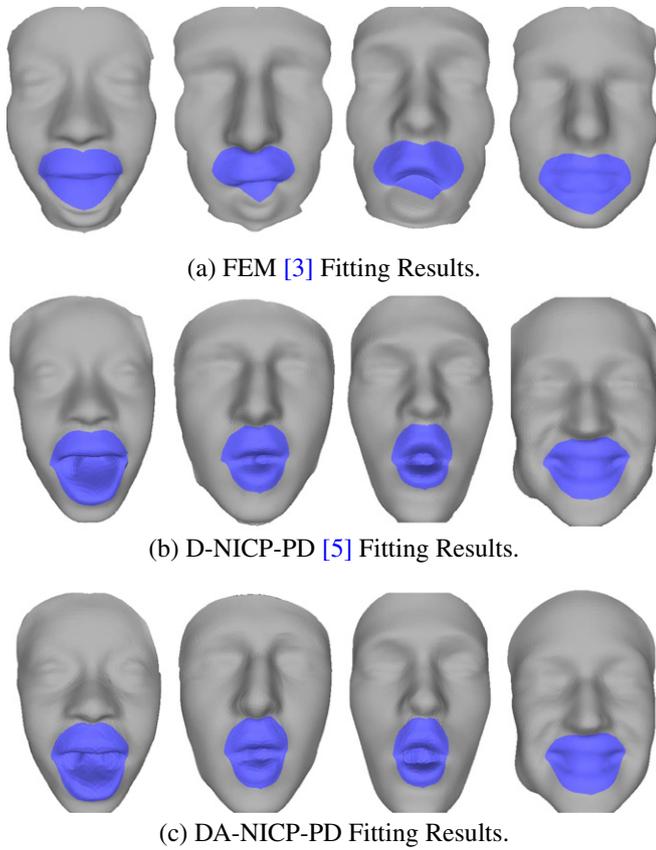


Fig. 3. Sample fitting results with annotated mouth region highlighted (in blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region. The second experiment is a facial expression recognition (FER) experiment on BU-4DFE database [47]. This experiment proves that an accurate fitting can benefit the recognition task (accurate dense alignment is important for accurate face and facial expression recognition [15–17]), even when a simple feature extraction and recognition method is used. In the last experiment, we evaluate the face recognition performance of the proposed active NICP method on the FRGC v2 database. This experiment aims at showing that our DA-NICP method can improve recognition performance, since quality 3D fittings is essential for face recognition [3,13].

In our experiments, we use the PCA models that are trained on the D-NICP-PD fitting results of the BU-3DFE database [48]. 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 models for the second subdivision level.

Table 2
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 shown in bold.

Method	Main face		Mouth	
	F_1 [%]	CR[%]	F_1 [%]	CR[%]
FEM [3]	75.65	75.87	73.49	73.38
D-NICP-PD [5]	79.24	79.47	78.18	78.13
D-NICP [5]	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

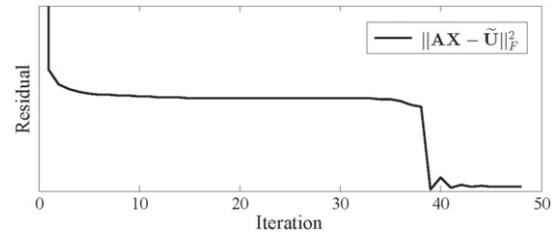


Fig. 4. Example residual error changes as the dynamic Active Non-rigid ICP (DA-NICP) fitting of the upper lip part in the second subdivision level progresses.

5.1. 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 [47] 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 1 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.

5.2. Facial expression recognition on BU-4DFE database

Next, we conduct facial expression recognition experiment on BU-4DFE database that consists of 4D faces (sequences of 3D faces), with temporal and spatial resolution being 25 frames/second and over 35,000 vertices, respectively. It includes 101 subjects each containing sequences of the six prototypical 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.

5.2.1. FER experiment protocol

We pick the apex frames for each sequence from temporal segment annotation provided in [16]. Due to the fact that each sequence has different apex period, the numbers of instances per expression differ from each other. An imbalanced training set would affect the quality of classifier, therefore, we trim the data for each apex sequence and generate a balanced set that consists of roughly 6000 instances, with each expression having nearly 1000 instances. Based on this data, we create a 10-fold partition, each time one fold is used for testing, all the others are used for training. Note that our BU-4DFE experiment is sequence-independent, which means the query sequence never appears in the training set.

Table 3
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 shown in bold.

Method (Coarse level fitting)	Main face		Mouth	
	F_1 [%]	CR[%]	F_1 [%]	CR[%]
D-NICP-PD [5]	78.31	78.53	77.23	77.23
D-NICP [5]	75.78	76.07	74.03	74.56
DA-NICP-PD	79.24	79.62	78.93	78.87
DA-NICP	76.39	76.61	75.67	75.99



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. Additionally, the 33 anchor points (as defined in Section 3.1) are plotted on the 3D mesh.

5.2.2. 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 divide the extracted region into non-overlapping blocks, for which HONV (Histogram of Oriented Normal Vectors) [49] features are computed (to this end other simple histogram-based

Table 4
Recognition rate (RR) achieved with FEM and DA-NICP method in FRGC v2 (%). The best performance achieved for each measure is shown in bold.

Methods	FEM [3]	DA-NICP
RR [%]	0.95	0.97

features can be used [50]). The final feature descriptor is the concatenation of feature vectors from the corresponding blocks.

Since the goal of our experiment is to compare 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 [51]. Radial Basis Function (RBF) kernel is employed as the distance metric. An empirical grid search is performed over the parameters in RBF kernel.

5.2.3. Experiment result and discussion

Table 2 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 drops 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. As can be seen, the residual error monotonically decreases and gradually converges to a minimum value. On average, the fitting of one facial part takes around 50 iterations.

To show the advantage of propagating from the second level (i.e. coarse level) to a third level in the hierarchy of subdivision surfaces, we conducted FER experiment using the coarse level fitting results from our proposed methods. Table 3 displays the FER performance of each method with the second level fitting. As can be seen from Tables 2 and 3, improvement from the second to the third level fitting is significant, with the increase varying from 0.5% to 1.5% in classification rate. Taking the proposed **DA-NICP-PD** method as an example, the classification rate rises from 79.62% to 81.16% (using main face feature), and rises from 78.87% to 80.41% with only mouth region feature. This can be easily understood since more useful facial details would be captured in the fine level fitting, and this should boost the performance of expression recognition.

5.2.4. 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. A typical example can be found in surprise example (row 6, column 5) of D-NICP-PD, where there are black holes around the mouth, this suggests inconsistency between mouth and neighbouring parts.

5.3. Facial recognition on FRGC v2 database

In this experiment, we want to demonstrate that the proposed Active Non-rigid ICP algorithm is not only capable of modelling facial expressions, but also capable of capturing important facial details for 3D face recognition. We use the FRGC v2 3D face database [52] which contains 3D face scans acquired using a Minolta 910 laser scanner that produces range images with a resolution of 640×480 in pixels. Here, we randomly select 200 neutral faces from 100 persons and separate them into two sets (i.e. train and test set), with each set containing one image from all 100 persons. We fit all the images using both FEM and the proposed active method (**DA-NICP**), and perform face recognition on the test set. In [17], the authors show that the normal azimuth angles combined with Principal Component Analysis (PCA) using a cosine-based distance measure can be used for robust face recognition from facial surfaces. Motivated by this work, we compute the azimuth angles of vertex normals and choose the absolute angle difference as the distance metric. Table 4 shows the recognition rates of both methods. Comparing with the 95% obtained by FEM, the proposed DA-NICP method achieves a recognition rate of 97%, which is a very strong evidence that a better modelling of 3D facial details leads to a better performance in the 3D face recognition task.

6. 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 on face expression recognition and face identification verify our theoretical expositions.

Acknowledgements

The work of S. Cheng and I. Marras is funded by the EPSRC project EP/J017787/1 (4D-FAB). M. Pantic acknowledges support by the European Community Horizon 2020 [H2020/2014-2020] under grant

agreement no. 645094 (SEWA). S. Zafeiriou also acknowledges support from EPSRC project EP/L026813/1 Adaptive Facial Deformable Models for Tracking (ADAManT).

References

- [1] V. Blanz, T. Vetter, A Morphable Model for the Synthesis of 3D Faces, Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '99., ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1999, pp. 187–194. <http://dx.doi.org/10.1145/311535.311556>.
- [2] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D Face Model for Pose and Illumination Invariant Face Recognition, Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on, 2009, pp. 296–301. <http://dx.doi.org/10.1109/AVSS.2009.58>.
- [3] G. Passalis, I. Kakadiaris, T. Theoharis, Intra-class retrieval of nonrigid 3D objects: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 29 (2) (2007) 218–229. <http://dx.doi.org/10.1109/TPAMI.2007.37>.
- [4] Y. Wang, G. Pan, Z. Wu, 3D Face Recognition in the Presence of Expression: A Guidance-based Constraint Deformation Approach, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–7. <http://dx.doi.org/10.1109/CVPR.2007.383277>.
- [5] B. Amberg, S. Romdhani, T. Vetter, Optimal Step Nonrigid ICP Algorithms for Surface Registration, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8. <http://dx.doi.org/10.1109/CVPR.2007.383165>.
- [6] B. Allen, B. Curless, Z. Popović, The space of human body shapes: reconstruction and parameterization from range scans, ACM Trans. Graph. (TOG) 22 (3) (2003) 587–594. <http://doi.acm.org/10.1145/882262.882311>. <http://dx.doi.org/10.1145/882262.882311>.
- [7] G. Pan, X. Zhang, Y. Wang, Z. Hu, X. Zheng, Z. Wu, Establishing point correspondence of 3D faces via sparse facial deformable model, IEEE Trans. Image Process. (TIP) 22 (11) (2013) 4170–4181. <http://dx.doi.org/10.1109/TIP.2013.2271115>.
- [8] D. Schneider, P. Eisert, Fast nonrigid mesh registration with a data-driven deformation prior, IEEE International Conference on Computer Vision Workshops (ICCV - W), 2009, pp. 304–311. <http://dx.doi.org/10.1109/ICCVW.2009.5457684>.
- [9] G. Tam, Z.-Q. Cheng, Y.-K. Lai, F. Langbein, Y. Liu, D. Marshall, R. Martin, X.-F. Sun, P. Rosin, Registration of 3D point clouds and meshes: a survey from rigid to nonrigid, IEEE Trans. Vis. Comput. Graph. 19 (7) (2013) 1199–1217. <http://dx.doi.org/10.1109/TVCG.2012.310>.
- [10] B. Amberg, R. Knothe, T. Vetter, Expression invariant 3D face recognition with a Morphable Model, Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2008, pp. 1–6. <http://dx.doi.org/10.1109/AFGR.2008.4813376>.
- [11] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 25 (9) (2003) 1063–1074. <http://dx.doi.org/10.1109/TPAMI.2003.1227983>. <http://dx.doi.org/10.1109/TPAMI.2003.1227983>.
- [12] X. Zhu, J. Yan, D. Yi, Z. Lei, S.Z. Li, Discriminative 3D morphable model fitting, Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG), IEEE, 2015, pp. 1–8. <http://dblp.uni-trier.de/db/conf/fg/fg2015.html#ZhuYLL15>.
- [13] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 29 (4) (2007) 640–649. <http://dx.doi.org/10.1109/TPAMI.2007.1017>.
- [14] V. Blanz, G. Basso, T. Poggio, T. Vetter, Reanimating faces in images and video, Comput. Graphics Forum 22 (3) (2003) 641–650. <http://dx.doi.org/10.1111/1467-8659.t01-1-00712>. <http://dx.doi.org/10.1111/1467-8659.t01-1-00712>.
- [15] G. Rajamanoharan, S. Zafeiriou, M. Pantic, L. Yin, Static and dynamic 3D facial expression recognition: a comprehensive survey, Image Vis. Comput. (IVC) 30 (10) (2012) 683–697. <http://www.sciencedirect.com/science/article/pii/S0262885612000935>.
- [16] G. Rajamanoharan, S. Zafeiriou, M. Pantic, D. Rueckert, Recognition of 3D facial expression dynamics, Image Vis. Comput. (IVC) 30 (10) (2012) 762–773. 3D Facial Behaviour Analysis and Understanding. <http://www.sciencedirect.com/science/article/pii/S0262885612000157>.
- [17] I. Marras, S. Zafeiriou, G. Tzimiropoulos, Robust Learning from Normals for 3D Face Recognition, Computer Vision - European Conference on Computer Vision Workshops and Demonstrations (ECCV-W), Lecture Notes in Computer Science > 7584, Springer Berlin Heidelberg, 2012, pp. 230–239. http://dx.doi.org/10.1007/978-3-642-33868-7_23. http://dx.doi.org/10.1007/978-3-642-33868-7_23.
- [18] I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, N. Murtuza, Multimodal face recognition: combination of geometry with physiological information, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2, 2005, pp. 1022–1029. <http://dx.doi.org/10.1109/CVPR.2005.241>.
- [19] d4d - 4D Capture Systems, 2011, <http://www.di3d.com/products/4dsystems/>.
- [20] A. Athana, S. Zafeiriou, S. Cheng, M. Pantic, Robust Discriminative Response Map Fitting with Constrained Local Models, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3444–3451. <http://dx.doi.org/10.1109/CVPR.2013.442>.
- [21] H. Li, R.W. Sumner, M. Pauly, Global Correspondence Optimization for Non-rigid Registration of Depth Scans, Proceedings of the Symposium on Geometry Processing, SGP '08., Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2008, pp. 1421–1430. <http://dl.acm.org/citation.cfm?id=1731309.1731326>.
- [22] T.B.A. Brunton, S. Wuhrer, Multilinear Wavelets: A Statistical Shape Space for Human Faces, 2014, CoRR abs/1401.2818. <http://arxiv.org/abs/1401.2818>.
- [23] Z. Wang, M. Grochulla, T. Thormahlen, H.-P. Seidel, 3D Face Template Registration Using Normal Maps, International Conference on 3D Vision (3DV), 2013, pp. 295–302. <http://dx.doi.org/10.1109/3DV.2013.46>.
- [24] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, N. Murtuza, Evaluation of 3D Face Recognition in the presence of facial expressions: an Annotated Deformable Model approach, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W), 2005, pp. 171. <http://dx.doi.org/10.1109/CVPR.2005.573>.
- [25] C. Mandal, H. Qin, B.C. Vemuri, A Novel FEM-based Dynamic Framework for Subdivision Surfaces, Proceedings of the Fifth ACM Symposium on Solid Modeling and Applications, SMA '99, ACM, New York, NY, USA, 1999, pp. 191–202. <http://doi.acm.org/10.1145/304012.304031>. <http://dx.doi.org/10.1145/304012.304031>.
- [26] G. Passalis, P. Perakis, T. Theoharis, I. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 33 (10) (2011) 1938–1951. <http://dx.doi.org/10.1109/TPAMI.2011.49>.
- [27] T. Fang, X. Zhao, O. Ocegueda, S.K. Shah, I. Kakadiaris, 3D/4D facial expression analysis: an advanced annotated face model approach, Image Vis. Comput. (IVC) 30 (10) (2012) 738–749.
- [28] D. Hahnel, S. Thrun, W. Burgard, An Extension of the ICP Algorithm for Modeling Nonrigid Objects with Mobile Robots, Proceedings of the 18th International Joint Conference on Artificial Intelligence, IJCAI'03, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003, pp. 915–920. <http://dl.acm.org/citation.cfm?id=1630659.1630791>.
- [29] T. Albrecht, M. Luthi, T. Vetter, A statistical deformation prior for non-rigid image and shape registration, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8. <http://dx.doi.org/10.1109/CVPR.2008.4587394>.
- [30] C.D. Schneider, P. Eisert, Algorithms for automatic and robust registration of 3D head scans, JVRB - J. Virtual Reality Broadcast. 7 (7). (2010) <http://nbn-resolving.de/urn:nbn:de:0009-6-26626>.
- [31] J.R. Tena, F. De la Torre, I. Matthews, Interactive Region-based Linear 3D Face Models, ACM SIGGRAPH New York, NY, USA, 2011, pp. 76:1–76:10. <http://doi.acm.org/10.1145/1964921.1964971>. <http://dx.doi.org/10.1145/1964921.1964971>.
- [32] T. Bolkart, S. Wuhrer, Statistical Analysis of 3D Faces in Motion, 3D Vision - 3DV 2013, 2013 International Conference on, 2013, pp. 103–110. <http://dx.doi.org/10.1109/3DV.2013.22>.
- [33] A. Salazar, S. Wuhrer, C. Shu, F. Prieto, Fully automatic expression-invariant face correspondence, Machine Vision and Applications 25 (4) (2014) 859–879. <http://dx.doi.org/10.1007/s00138-013-0579-9>. <http://dx.doi.org/10.1007/s00138-013-0579-9>.
- [34] A. Athana, S. Zafeiriou, G. Tzimiropoulos, S. Cheng, M. Pantic, From pixels to response maps: discriminative image filtering for face alignment in the wild, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 37 (6) (2015) 1312–1320. <http://dx.doi.org/10.1109/TPAMI.2014.2362142>.
- [35] A. Athana, S. Zafeiriou, S. Cheng, M. Pantic, Incremental Face Alignment in the Wild, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1859–1866. <http://dx.doi.org/10.1109/CVPR.2014.240>.
- [36] L. Dutreue, A. Meyer, S. Bouakaz, Feature Points Based Facial Animation Retargeting, Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology, VRST '08, ACM, New York, NY, USA, 2008, pp. 197–200. <http://doi.acm.org/10.1145/1450579.1450621>. <http://dx.doi.org/10.1145/1450579.1450621>.
- [37] T.C.S. Rendall, C.B. Allen, Reduced surface point selection options for efficient mesh deformation using radial basis functions, J. Comput. Phys. 229 (8) (2010) 2810–2820. <http://dx.doi.org/10.1016/j.jcp.2009.12.006>. <http://dx.doi.org/10.1016/j.jcp.2009.12.006>.
- [38] A. Jacobson, I. Baran, J. Popović, O. Sorkine, Bounded Biharmonic Weights for Real-time Deformation, ACM SIGGRAPH, ACM, New York, NY, USA, 2011, pp. 78:1–78:8. <http://doi.acm.org/10.1145/1964921.1964973>. <http://dx.doi.org/10.1145/1964921.1964973>.
- [39] A. Johnson, Spin-Images: A Representation for 3-D Surface Matching, (Ph.D. thesis) Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1997,
- [40] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 14 (2) (1992) 239–256. <http://dx.doi.org/10.1109/34.121791>.
- [41] D. Chetverikov, D. Stepanov, P. Krsek, Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm, Image Vis. Comput. (IVC) 23 (2005) 299–309.
- [42] P. Siarry, G. Berthiau, F. Durdin, J. Haussy, Enhanced simulated annealing for globally minimizing functions of many-continuous variables, ACM Trans. Math. Softw. (TOMS) 23 (2) (1997) 209–228. <http://doi.acm.org/10.1145/264029.264043>. <http://dx.doi.org/10.1145/264029.264043>.
- [43] G. Papaioannou, E.A. Karabassi, T. Theoharis, Reconstruction of three-dimensional objects through matching of their parts, IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) 24 (1) (2002) 114–124. <http://dx.doi.org/10.1109/34.982888>.

- [44] C. Loop, *Smooth Subdivision Surfaces Based on Triangles*, Master's Thesis, The University of Utah, 1987.
- [45] J. Warren, H. Weimer, *Subdivision Methods for Geometric Design: A Constructive Approach*, first ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001.
- [46] E. Catmull, J. Clark, *Seminal Graphics*, Ch. Recursively Generated B-spline Surfaces on Arbitrary Topological Meshes ACM, New York, NY, USA, 1998, pp. 183–188. <http://doi.acm.org/10.1145/280811.280992>. <http://dx.doi.org/10.1145/280811.280992>.
- [47] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3D dynamic facial expression database, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2008. pp. 1–6. <http://dx.doi.org/10.1109/AFGR.2008.4813324>.
- [48] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2006. pp. 211–216. <http://dx.doi.org/10.1109/FGR.2006.6>.
- [49] S. Tang, X. Wang, X. Lv, T.X. Han, J. Keller, Z. He, M. Skubic, S. Lao, Histogram of Oriented Normal Vectors for Object Recognition with a Depth Sensor, *IEEE Asian Conference on Computer Vision (ACCV)*, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 525–538. http://dx.doi.org/10.1007/978-3-642-37444-9_41. http://dx.doi.org/10.1007/978-3-642-37444-9_41.
- [50] G. Rajamanoharan, S. Zafeiriou, M. Pantic, Binary Pattern Analysis for 3D Facial Action Unit Detection, *Proceedings of the British Machine Vision Conference (BMVC)*, Guildford, UK, 2012.
- [51] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, *ACM transaction on intelligence system technology* 2 (3) (2011) 27:1–27:27. <http://doi.acm.org/10.1145/1961189.1961199>. <http://dx.doi.org/10.1145/1961189.1961199>.
- [52] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Washington, DC, USA, 2005, pp. 947–954. <http://dx.doi.org/10.1109/CVPR.2005.268>. <http://dx.doi.org/10.1109/CVPR.2005.268>.