
Proceedings Paper:
Pears, N. E. orcid.org/0000-0001-9513-5634, Dai, Hang, Smith, William Alfred Peter orcid.org/0000-0002-6047-0413 et al. (1 more author) (2023) Laplacian ICP for Progressive Registration of 3D Human Head Meshes. In: International Conference on Automatic Face and Gesture Recognition 2023. International Conference on Automatic Face and Gesture Recognition 2023, 05-08 Jan 2023 IEEE , USA

Laplacian ICP for Progressive Registration of 3D Human Head Meshes

Anonymous FG2023 submission
Paper ID 68

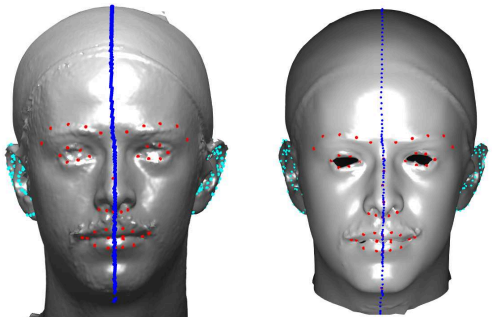
***Abstract*— We present a progressive 3D registration framework that is a highly-efficient variant of fully-automatic, classical non-rigid ICP. Since it uses the Laplace-Beltrami operator for deformation regularisation, we view the overall process as Laplacian ICP (L-ICP). This exploits a 'small deformation per iteration' assumption and is progressively coarse-to-fine, employing an increasingly flexible deformation model, an increasing number of correspondence sets, and increasingly sophisticated correspondence estimation. Correspondence matching is only permitted within predefined vertex subsets derived from domain-specific feature extractors. Additionally, we present a new benchmark and a pair of evaluation metrics for 3D non-rigid registration, based on annotation transfer. We use this to evaluate our framework on a publicly-available dataset of 3D human head scans (Headspace). The method is highly robust and only requires a small fraction of the computation time compared to the most popular classical approach, yet has comparable registration performance.**

I. Introduction

Determining surface correspondences across a set of 3D shapes is key to modelling them. One approach employs non-rigid transformation of a template (source) shape, so that its vertices align with those of a target shape - see Figure 1. When a template is non-rigidly registered to a set of shapes of some class, this enables construction of statistical shape models, such as 3D Morphable Models (3DMMs, [1], [23]).

Non-rigid registration has been extensively explored both in terms of classical optimisation algorithms and deep learning. Often, the latter requires a large corpus of training data, data augmentation techniques or transfer learning. Here, we revisit the classical approaches, which do not have such requirements and are of high utility in low data volume cases. In this respect, we provide a new formulation of dense, non-rigid Iterative Closest Points.

In essence, our algorithm incorporates a form of progressively-relaxed Laplacian deformation regularisation into an progressively coarse-to-fine ICP-style framework [3], [1] - and hence our approach is termed Laplacian ICP (L-ICP). Laplacian mesh editing [25] is known to be computationally efficient, due to its sparse linear structure and hence a key benefit of L-ICP is that it is very efficient computationally, when compared with competing classical approaches for non-rigid registration, such as Optimal-step N-ICP [1] and Coherent Point Drift (CPD [21]). The approach is designed to incorporate shape morphing constraints supplied by domain-specific feature extraction algorithms, as shown in Figure 1, which has facial landmarks, ear landmarks and an intrinsic symmetry contour on a human head. Furthermore, it handles variable mesh resolution via the cotangent weighting scheme of the Laplace-Beltrami operator.



Target Data Template Morph

Fig. 1: Target scan (left) and template (right) morphed with Laplacian ICP. Correspondence sets are: (i) landmarks (red); (ii) right ear landmarks (cyan); (iii) left ear landmarks (cyan); (iv) symmetry contour (blue); and (v) all remaining vertices on mesh (grey surface).

L-ICP is packaged in a very flexible, *staged* non-rigid registration framework, where stages are defined by high-level scripting. Therefore, it is easily adapted to different shape classes, guided by domain-specific sets of automatically-extracted correspondences. It is a progressively coarse-to-fine process that, as it transitions into each new stage, may employ: i) an increasing number of correspondence sets (landmarks, contours, regions); ii) an increasingly refined correspondence estimation; iii) an increasingly flexible shape deformation model and iv) an iterative shape refinement, used after correspondences stabilise, that generates a Laplace-Beltrami operator consistent with the template itself. Since our algorithm estimates the deformed shape directly in a linear system, it is easy to incorporate constraints from other linear relations on the deformation, such as those from 3DMMs or extrinsic symmetry constraints. We evaluate our work on the Headspace dataset of 3D human head scans [9] and compare to the per-vertex affine regularisation approach from the most commonly-used N-ICP variant [1].

In summary, our contributions are: i) fully-dense morphing via progressively relaxed Laplace-Beltrami regularisation; ii) a flexible and progressive coarse-to-fine registration framework; iii) a new publicly-available non-rigid registration benchmark for the Headspace [9] human head dataset, comprised of a set of manual annotations (859 subjects) and a pair of annotation transfer metrics. We will make our annotation data and registration code available in the interests of reproducibility.

II. Related work

Non-rigid 3D registration, correspondence matching and 3DMM fitting are highly-active research areas. Methods focus on correspondences across organic shape sets such as faces [5], [17], [9], [23], human bodies [4], [20], [13], [24] and various human organs [15]. Alternatively, they focus on man made objects, such as chairs, cups and aircraft [28], [19]. A current popular approach, due to its compactness and flexibility, is to employ implicit surface representations; for example, where the 3D surface is the zero level set of a learnt Signed Distance (to surface) Function (SDF) [22], [11], or where the function returns an occupancy indicator. Zheng *et al*. [30] propose the use of Deep Implicit Templates for 3D shape representation. Their method employs a Deep Implicit Function (DIF) that is decomposed into a warping function and an implicit template. The warping function transforms point samples of the raw scan to their canonical positions, which are then mapped to SDF values by the implicit template. In contrast to [12], their approach is able to learn an implicit template for some shape class. The method also establishes dense correspondences across the shape dataset in an unsupervised fashion.

Rather than the standard pipeline of multi-view 3D shape reconstruction, followed by a 3D-to-3D template to data registration scheme, the ToFu system of Li *et al*. [18] outputs a 3D facial mesh in a consistent topology directly from a set of calibrated 2D input images. The idea here is to consider a feature volume, where a grid of points in that volume selects 2D features by projection into the multi-view images. A coarse-to-fine architecture is then able to predict a consistent facial mesh topology.

In this work, we revisit classical ICP [3] in its non-rigid form [1]. Widely-used methods of classical non-rigid registration include Non-rigid Iterative Closest Point (NICP) [1], Coherent Point Drift (CPD) [21], Thin Plate Spline (TPS) approaches [6], [29] and the method of Li *et al*. method [16], which employs a Levenberg-Marquardt based optimisation. The *As-Rigid-As-Possible* (ARAP) form of deformation regularisation was introduced by Sorkine *et al*. [25] in their work on surface editing. Here the mesh Laplacian was employed to ensure that shape details are preserved. This proved extremely effective, as the relative location of vertices and hence local shape is encoded by the Laplacian. Furthermore, it was noted that by using the cotangent scheme in the computation of the Laplacian operator, non-uniformity in 3D mesh resolution and connectivity could be accommodated. Dai *et al*. [8] use the Laplace-Beltrami shape regularisation as a way of initialising the Coherent Point Drift [21] algorithm. Although we employ a similar early-stage template adaptation, [8] use fixed landmarks with a single fixed stiffness weight, whereas ours uses both fixed and variable correspondence sets and a stiffness weighting schedule. More importantly, we demonstrate, for the first time, that it is possible to use LB regularisation to do fast, fully dense, and complete shape morphing.

III. Laplacian ICP

Any form of non-rigid 3D ICP seeks to iteratively refine a source shape, such that it becomes registered with a target shape. A key question is how to regularise the deformation of the source mesh. Inspired by Laplacian surface editing [25], we employ the Laplace-Beltrami (LB) operator in our regularisation term. When the LB operator is applied to a mesh, it extracts vectors in the direction of the local surface normal, with magnitude proportional to the local mean surface curvature. This computation is of a discrete form and here we employ the cotangent approximation scheme. Thus, within our iterative optimisation scheme, we consider a regularising energy term, E_{reg}, of the form

$$E_{reg}(\mathbf{X}_{i+1}) = ||\mathbf{L}_{i+1}\mathbf{X}_{i+1} - \mathbf{L}_i\mathbf{X}_i||_F^2 \quad (1)$$

where $\mathbf{X} \in \mathbb{R}^{N\times 3}$ is a matrix of N source mesh vertex positions, $\mathbf{L}_i = L(\mathbf{X}_i)$ is the LB operator ($\mathbf{L} \in \mathbb{R}^{N\times N}$, sparse) computed from the source shape (*e.g*. template) at the ith iteration and $||.||_F$ is the Frobenius norm. This regularisation simultaneously applies an orientation constraint on the template update, because of the extracted surface normal directions - and a shape constraint, due to the surface normal magnitudes being proportional to local mean surface curvature. Note that $\mathrm{rank}(\mathbf{L}) = N - 1$. A physical interpretation of this is that a pure translation applied to all vertices of $\mathbf{X}$ would provide no change in the energy described by Eqn. 1. The positional error associated with at least one pair of corresponding vertices from template to data can provide the necessary additional constraint as a shape error, E_{shp}, that we aim to minimise. We form a weighted combination of our energy terms as:

$$E = E_{shp} + \lambda E_{reg}, \quad (2)$$

where the parameter λ balances the influence of the two component energies. Specifically, the energy for a new deformation is given as:

$$E(\mathbf{X}_{i+1}) = ||\mathbf{P}_i\mathbf{X}_{i+1} - \mathbf{Q}_i\mathbf{Y}_i||_F^2 + \lambda_i E_{reg}(\mathbf{X}_{i+1}) \quad (3)$$

where $\mathbf{Y}_i$ are the target data vertices, $\mathbf{P}_i$ and $\mathbf{Q}_i$ are highly-structured binary selection matrices (discussed later) that define source-target bijective correspondences and λ_i is a weighting that defines the amount of mesh deformation regularisation.

Clearly, Eqn. 3 is not closed form, as the regularisation term, $E_{reg}(.)$, at the update step $(i+1)$, is dependent on $\mathbf{L}_{i+1}$, which is a function of $\mathbf{X}_{i+1}$, *i.e*. the updated template itself. This suggests an iterative procedure, where we invoke a *small deformation* per-iteration assumption, so that $\mathbf{X}_{i+1} \approx \mathbf{X}_i$, which implies $\mathbf{L}_{i+1} \approx \mathbf{L}_i$ and that we can compute an accurate regularising term, E_{reg}. We ensure this small deformation assumption holds by initialising L-ICP with a high value of λ_i giving a large template mesh 'stiffness'. This parameter gradually becomes smaller as the template gets closer to the target data, thus balancing it with a smaller E_{shp} term. In other words, we employ a regularising mesh stiffness schedule, of gradually decreasing stiffness. This is

analagous to the mesh stiffness schedule defined in N-ICP [1] that regulates mesh deformation by limiting differences in locally-affine transformations. Thus λ changes in each iteration of L-ICP, as do the data vertices $\mathbf{Y}$, as any rigid component of template-data alignment is applied to the data rather than the template, for reasons described later.

We use a set of domain-specific feature extractors to generate C subsets of source vertices (selected by $\mathbf{P}^{1\ldots C}$) and target vertices (selected by $\mathbf{Q}^{1\ldots C}$) that are in correspondence with each other. Using these correspondence sets and the small deformation approximation for regularisation, we have our overall energy term as:

$$E_{i+1} = \sum_{j=1}^{C} \alpha_j ||\mathbf{P}_i^j \mathbf{P}^j \mathbf{X}_{i+1} - \mathbf{Q}_i^j \mathbf{Q}^j \mathbf{Y}_i||_F^2 + \lambda_i ||\mathbf{L}_i(\mathbf{X}_{i+1} - \mathbf{X}_i)||_F^2, \tag{4}$$

where $(\mathbf{P}_i^j, \mathbf{Q}_i^j)$ select correspondences from within correspondence sets $(\mathbf{P}^j, \mathbf{Q}^j)$. Note that $E_{i+1} = E(\mathbf{X}_{i+1})$ and α_j is a weighting that expresses the relative confidence in the jth correspondence set. Eqn. 4 requires a linear solve for $3N$ variables per iteration, whereas N-ICP [1] requires a linear solve for $12N$ variables as per-vertex affine transformations are computed. Furthermore, the L-ICP constraint is vertex-based, whereas the N-ICP constraint is edge-based, giving around three times as many shape regulation equations in the linear solve for a triangular mesh. As a result, L-ICP is much more compact and efficient than N-ICP.

A. *Fixed correspondence shape refinement*

Suppose that, within some iteration, i, of L-ICP, we fix both the correspondences (all C $\mathbf{P}_i, \mathbf{Q}_i$ matrices) and the parameter λ_i and iteratively minimise the energy defined in Eqn. 4 by updating the template mesh $\mathbf{X}$ (*i.e.* employ an inner optimisation loop). This drives the regularisation term, E_{reg}, to zero, allowing the template to move closer to the data in steps of decreasing size until the recomputed template, $\mathbf{X}_{i+1}$, and the LB operator employed in the update, $\mathbf{L}_i$, become consistent with each other. In practice, we find that this second-order template deformation only takes a few iterations until the change in $\mathbf{X}$ over an iteration becomes small. For some fixed set of correspondences, this process provides a small template shape refinement. Therefore, it is only used in the final morphing stage of our framework, when the set of correspondences has become stable.

IV. Coarse-to-fine L-ICP framework

Our L-ICP registration framework is shown in Fig. 2. This defines a set of $s = 1 \ldots S$ user-defined, application-dependent, coarse-to-fine registration stages, each of which terminates when the stage's template deformation falls below some threshold, $||\Delta\mathbf{X}||_F^2 < t_s$, or a maximum number of iterations, i_s^{max} is reached. Within each stage, the user defines: i) a set of correspondence sets; ii) a correspondence strategy; iii) a template deformation model, and iv) stage termination conditions, (t_s, i_s^{max}). The end user can rapidly define a morphing process as a set of such stages, as each stage inherits properties from the previous stage, unless they are re-specified. Typically, early registration stages have a few landmark-based correspondence sets, coarse correspondence matching and low-dimensional deformation models (*e.g.* global affine or low-dimensional 3DMMs). Later stages have many correspondence sets, fine correspondence matching and high-dimensional deformation models (*e.g.* high-dimensional 3DMMs or free movement with $3N$ degrees of freedom). A key feature is that we switch to a higher-dimensional deformation model before switching to dense correspondence search over the full surface. With otherwise free movement of the template vertices, this relies on the aforementioned Laplacian shape regularisation to effectively interpolate between the sparse correspondences, thus adapting the source template to the target shape. In essence, such *template adaptation* is akin to a fully-automatic, iterative mesh editing process and is the reason that our method is successful on widely different target shapes using a single template (*e.g.* small babies heads and large adult heads). Our framework is very flexible and is designed to exploit situations where there are landmarks and/or contours and/or surface regions (*e.g.* from semantic parts segmentation) that correspond across the source and target shapes. In the case of landmarks, a bijective (one-to-one) correspondence is predefined, and for contours and regions, we employ mutual nearest neighbour search. We now detail the stage-selectable choices in our L-ICP framework that relate to Fig 2.

A. *Correspondence sets for the human head*

Our example application is to register the FaceWarehouse head template [7] with the Headspace dataset of 3D human heads [9]. In order to generate the correspondence selection matrices, $\mathbf{P}_i^j, \mathbf{Q}_i^j$ in Eq. 4, which define sets of corresponding mesh vertices, we employ a 3D face landmarker system, a 3D ear landmarker system, our own symmetry contour extractor, and a large correspondence set region that is all vertices that are otherwise unused in these landmark and contour sets. The face and ear landmarkers employ the 2D channel as well as the 3D, whereas the symmetry contour extraction is based on 3D data only.

The face landmarker system uses the standard `dlib` face landmarker, which is based on the work of Kazemi and Sullivan [14]. This extracts 68 2D facial landmarks of which we retain 52, discarding the 16 that follow the apparent contour of the face. We project these 2D facial landmarks to their nearest vertices on the target 3D data scan.

We employ the *Human Ear Reconstruction Autoencoder* (HERA) system of Sun *et al.* [27] to generate 55 landmarks per ear. This regresses the pose and shape parameters of a 3D Morphable Model (3DMM) of the human ear [10], such that a synthesised 2D image of the ear matches a rendered image of the colour-textured 3D target data. Two side views of the target 3D head can be determined using the facial landmarks, and the left ear is reflected enabling us to use a single right ear model. Again, 2D landmarks are projected to their nearest 3D vertices, and the residual Euclidean distance is stored, allowing the weighting of individual landmarks (larger distances have lower weights).

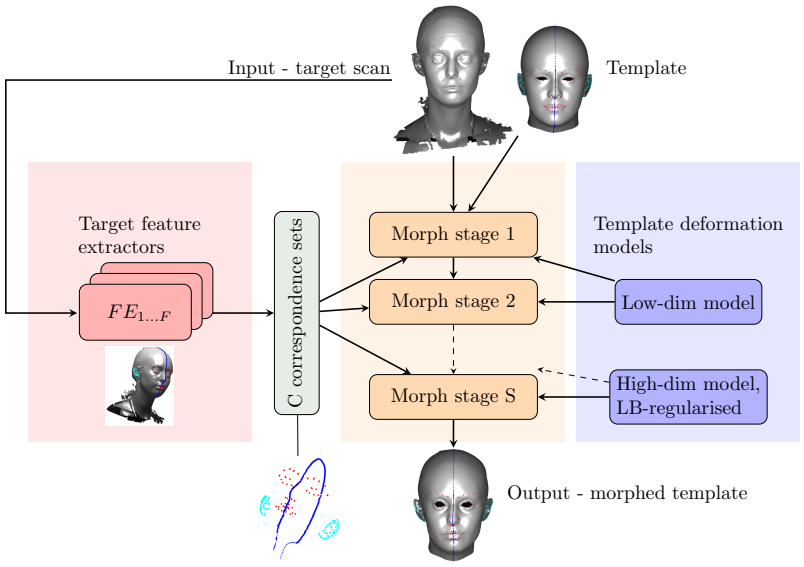


Fig. 2: L-ICP registration framework. Central orange panel: coarse-to-fine registration stages. Left pink panel: feature extractors for target data. Right purple panel: template deformation models.

We incorporate the method of Benz *et al*. [2] for symmetry plane extraction into a more general procedure for symmetry contour generation. The piecewise nature of our algorithm allows the extracted symmetry contour to be intrinsic; for example, if the nose is bent to one side, it successfully tracks the nose ridge.

Finally the sets of template/data vertices, $\mathbb{B}_{t,d}$, not designated as face/ear landmarks and not on the symmetry contour are defined for region-based correspondences *e.g.* for the template: $\mathbb{B}_t = \{\mathbb{X} \setminus \{\mathbb{F}_t \cup \mathbb{E}_t^l \cup \mathbb{E}_t^r \cup \mathbb{S}_t\}\}$, where $\mathbb{X}$ is the set of all vertices on the templates and $\mathbb{F}_t, \mathbb{E}_t^l, \mathbb{E}_t^r, \mathbb{S}_t$ are the sets of template face landmarks, left ear landmarks, right ear landmarks and symmetry contour on the template respectively.

B. Correspondence matching

We perform a correspondence match for all active correspondence sets that do not consist of fixed landmarks. These relate to either corresponding surface contours or corresponding surface regions. Our framework enables one of three approaches to be selected. All three approaches are reliant on *mutual nearest neighbours* (MNN), where we take the subset of the bidirectional 1-nearest neighbour search results, such that the correspondence relation is bijective. Additiionally, this mutual nearest neighbour search can specified to be over the 3-DOF vertex positions or 6-DOF vertex positions and their associated normal vectors, with an weighting factor between positions and their normals. Firstly, MNN search in itself is a suitably conservative approach for early-stage morphing and has the benefits of handling mesh holes automatically and obviating the need for a manually-tuned threshold to filter out bad correspondences. Secondly, in the *normal shooting* method, the MNN approach is augmented by projecting a vector from the template to its corresponding data vertex along the source normal, which in general, results in an off-surface target point. However, when the source and target are close to each other, this is often a better morph direction, due to the generally different spatial sampling of the two surfaces.

C. Deformation models

Early stages of deformation use a global affine model, described next, whereas later stages use LB-regularised template deformation.

1) Affine deformation: Suppose that the template shape is given as a matrix of vertex positions, $\mathbf{X_i} \in \mathbb{R}^{N\times 3}$, after the $(i-1)th$ shape update with initial shape $\mathbf{X}_1$, and that the data, whose 6 DoF pose may vary, is given as the matrix of vertex positions $\mathbf{Y_i} \in \mathbb{R}^{N_Y\times 3}$, then we solve the following linear least squares equation for the affine transform $\mathbf{A_i} \in \mathbb{R}^{4\times 3}$:

$$\begin{bmatrix} \alpha_1 \mathbf{P}_i^1 \mathbf{P}^1 \mathbf{X_i} & \alpha_1 \mathbf{1}_i^1 \\ \vdots & \vdots \\ \alpha_C \mathbf{P}_i^C \mathbf{P}^C \mathbf{X_i} & \alpha_C \mathbf{1}_i^C \end{bmatrix} \mathbf{A}_i = \begin{bmatrix} \alpha_1 \mathbf{Q}_i^1 \mathbf{Q}^1 \mathbf{Y_i} \\ \vdots \\ \alpha_C \mathbf{Q}_i^C \mathbf{Q}^C \mathbf{Y_i} \end{bmatrix}, \quad (5)$$

where $(\alpha_1 \dots \alpha_C)$ are the relative influence weights for various sets of correspondences. $\mathbf{P}_i^j \in \{0,1\}^{N^j \times N_a^j}$ is a binary selection matrix that selects N^j vertices associated with the j*'th* correspondence set $(j = 1 \dots C)$ from the set of *all* N_a^j vertices in that correspondence set and $\mathbf{P}^j \in \{0,1\}^{N_a^j \times N}$ is the binary matrix that selects *all* members of the correspondence set from the template vertices. Note that if the correspondence set contains fixed landmarks, then $\mathbf{P}_i^j = \mathbf{I}_{N_a^j}$, otherwise it is determined by mutual nearest neighbour correspondence search, whereas $\mathbf{P}^j$ are constant matrices. Also note that $\mathbf{1}_i^j$ is a vector of 1s with length equal to the number of correspondences on iteration i for

correspondence set j. The matrices $\mathbf{Q}_i^j, \mathbf{Q}^j$ are binary selection matrices that select data vertices in an analagous way to the template selection matrices. We choose to decompose the affine transform into a rigid part and a non-rigid part, such that

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{R}_i\mathbf{B}_i \\ \mathbf{t}_i \end{bmatrix} \tag{6}$$

where $\mathbf{R}_i \in \mathbb{R}^{3\times3}$ is a rotation matrix, $\mathbf{B}_i \in \mathbb{R}^{3\times3}$ is a non-rigid deformation matrix composed of anisotropic scaling and shears and $\mathbf{t}_i \in \mathbb{R}^{1\times3}$ is a translation vector. We then apply the rigid part of the affine transform to the data:

$$\mathbf{Y}_{i+1} = (\mathbf{Y}_i - \mathbf{1}_{\mathbf{N_Y}}\mathbf{t}_i)\mathbf{R}_i^{\mathrm{T}} \tag{7}$$

where $\mathbf{1}_{\mathbf{N_Y}}$ is a column vector of N_Y 1s. The non-rigid part of the affine transform is applied to the template:

$$\mathbf{X}_{i+1} = \mathbf{X}_i\mathbf{B}_i \tag{8}$$

We could apply the full affine transform to the template, so this may seem like an unnecessary complication. However, it is very useful to employ this decomposition, which maintains a canonical pose of the template, particularly in variants of L-ICP that constrain template deformation to be symmetrical, or employ a 3DMM to reduce the dimensionality of the template deformation model.

2) LB regularised template deformation: In later stages of the morphing process, we wish to deform the template in a more detailed way that cannot be modelled by a simple low-dimensional transform. To achieve this, we solve directly for source mesh vertex positions, under the regularisation of the Laplace-Beltrami constraint. Specifically, we minimize the energy in Equation 4 by iteratively solving for $\mathbf{X}_{i+1}$ in the following weighted linear least-squares problem:

$$\begin{bmatrix} \alpha_1\mathbf{P}_i^1\mathbf{P}^1 \\ \vdots \\ \alpha_C\mathbf{P}_i^C\mathbf{P}^C \\ \lambda_i\mathbf{L_i} \end{bmatrix} \mathbf{X}_{i+1} = \begin{bmatrix} \alpha_1\mathbf{Q}_i^1\mathbf{Q}^1\mathbf{Y}_i \\ \vdots \\ \alpha_C\mathbf{Q}_i^C\mathbf{Q}^C\mathbf{Y}_i \\ \lambda_i\mathbf{L_i}\mathbf{X}_i \end{bmatrix}, \tag{9}$$

where α_j are relative influence weights for various sets of correspondences and λ_i is the mesh stiffness weight at iteration i of the deformation stage.

V. 3D Registration of the Human Head

Correspondence sets are weighted using empirical grid search as follows: face landmarks 1.5, symmetry contour 1.4, left/right ear landmarks 1.0, all other vertices 1.0. We define five stages within our framework for human head registration and an example of the morphing stage outputs is shown in Fig. 3.

Stage 1 - Affine template initialisation. Goal: align the data to the template, transform the template to the same scale and aspect ratio of the data. Settings: i) correspondence sets, $C = 3$: face landmarks, left ear landmarks, right ear landmarks; ii) correspondence matching: mutual nearest neighbours (MNN); iii) deformation model: global affine (one shot). We solve the linear least squares problem in Eq 5 for the required global affine deformation, which is then decomposed and distributed between the target data (rigid part) and the template (non-rigid part), as described in Eq 6 to 8.

Stage 2 - Affine template adaptation. Goal: improve depth and height scaling using symmetry contour. Settings: i) correspondence sets, $C = 4$: three landmark sets from previous stage plus the symmetry contour; ii) correspondence matching: MNN; iii) deformation model: global affine (iterative). We iteratively compute the affine template update, $\mathbf{A}_i$ using Eq 5 and perform the template/target updates according to Eq 6 to 8. A small number of iterations stabilises the symmetry contour correspondences, maximum number of iterations is set at 15.

Stage 3 - Laplacian template adaptation. Goal: adapt the template shape to the landmarks and symmetry contour. Settings: i) correspondence sets, $C = 4$, same as previous stage; ii) correspondence matching: MNN; iii) deformation model: LB-regulated free vertex deformation (iterative). This stage significantly deforms the source template, as it recomputes better correspondences along the symmetry contour, while simultaneously decreasing the regularising mesh stiffness (λ_i in Eq 9), according to some predefined schedule, at each iteration. The stiffness parameter ranges from 100 to 0.1 with a maximum number of iterations set at 58.

Stage 4 - Morphing with dense correspondences. Goal: compute dense correspondences for dense morphing. Settings: i) correspondence sets, $C = 5$: all landmark sets plus symmetry contour plus the 'set difference' region; ii) correspondence matching: MNN; iii) deformation model: LB-regulated free vertex deformation (iterative). Here we extend correspondences to the *set difference region* - i.e. the template region vertices that are not already used as either a landmark or on the symmetry contour. The stiffness parameter ranges from 100 to 1 with maximum iterations set to 31.

Stage 5 - Morphing with normal shooting. Goal: employ more refined dense correspondences for dense morphing. This stage re-specifies the correspondence search to the *normal shooting* correspondence method, but otherwise inherits all other framework selections from stage 4. The stiffness parameter ranges from 0.9 to 0.1 with maximum iterations set to 27. This means the maximum number of shape change iterations is 132 over all stages.

VI. Evaluation

Quantitative evaluation of 3D shape registration and correspondence quality using real-world data is notoriously difficult due to a lack of high-quality ground truth data. One approach is to use a proxy evaluation where better correspondences are deemed to be those that build better statistical models according to some metrics. For example, Styner *et al.* [26] proposed the use of three 3DMM metrics - compactness, generalisation and specificity. However, these are only meaningful when the template is on or very close to the data surface (results in supplementary). Furthermore, for soft organic shapes, strong perceptual consensus on what is a good correspondence often only exists at a very sparse

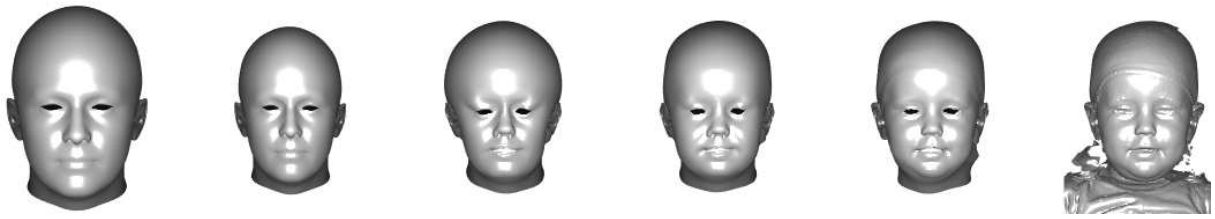


Fig. 3: Far left shows the employed Facewarehouse [7] full head template, far right shows the raw 3D Headspace data [9]. Intermediate images show morph stages 2-5 of our L-ICP framework. In most applications, only the final registration quality matters (column 5) and we see that the poor shape around the eyes after stage 3 (column 3) has been corrected. Zoom required, more morphs in supplementary.

set of surface locations. In our human head example, these are the physical junctions of tissues, such as eye and mouth corners. To mitigate these problems, we propose a different form of evaluation procedure and benchmark, which is based on the manual annotation of *facial contours*. Quantitative metrics are proposed that capture both the repeatability and homogeneity of how such annotations are transferred from the data onto the non-rigidly registered template. These are detailed further in Section VI-B, but we first describe the dataset used.

A. Dataset

To evaluate the L-ICP framework, we use the publicly-available Headspace dataset [9] of high resolution (150K-200K vertices) 3D images of the human head. The primary format for this dataset is OBJ, which provides vertex coordinates, mesh connectivity and texture coordinate information. It is also equipped with five 2D images per subject, used to texture map the 3D mesh surface using the supplied texture coordinates, and five corresponding camera calibration files with both intrinsic and extrinsic camera parameters. We employ two of the five views (left-frontal and right-frontal) to manually annotate a range of facial feature contours on the 2D images, including those around the eyes and eyelid, mouth, nose, nasolabial folds and ears. In the case of facial contours that are not well-defined in the image pair, annotators are instructed to omit them. Due to the labour-intensive nature of this, we have around 57% coverage of the Headspace dataset.

B. Annotation transfer

We transfer the 2D target data annotations to the 3D target mesh, which is achieved via left/right camera ray-to-mesh intersection, where rays are generated using the left/right camera calibration matrices. These 3D target scan mesh surface coordinates are then transferred on to their nearest neighbor template mesh vertices, after registration. If multiple annotations (*e.g.* left and right view of the same facial contour) transfer to the same 3D template coordinate, that is recorded as a single hit. Examples of this annotation transfer over a wide subject age range are given in Figure 4.

In the case of a consistent non-rigid registration performance, the annotations on different subject's target data scans should transfer to the same vertices on the template, or at least to closely neighboring vertices. Thus, by measuring the template surface density of such an *annotation transfer* process, we can generate a quantitative evaluation of registration repeatability. Additionally, this can be qualitatively evaluated by color-mapping the template with the density of those annotation transfers. These should be sharply defined on the template. Figure 5-left shows a front and side view of the employed Facewarehouse [7] template, with its surface color mapped with the frequency of the annotation transfers over $N_{subj} = 675$ Headspace subjects, with dark blue being zero transfers and yellow being the maximum frequency of transfer. The transfer is highly-repeatable around facial features, especially the mouth and nose. This is expected, as the registration process is well-guided by automatic landmarking in these regions. In contrast, there is lower repeatability around the nasolabial folds.

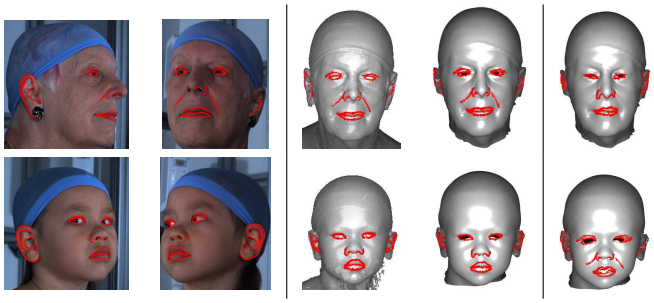


Fig. 4: Left two columns: images with manual annotations. Third column: two-view annotations amalgamated and projected to 3D target scans. Fourth column - morphed templates with annotations transferred. Fifth column - annotations swapped across the two subjects - diverse annotation swaps in supplementary.

1) Annotation transfer metrics: A quantitative repeatability metric should indicate high performance when template vertices are selected in the transfer numerous times (in the best case, N_{subj} times) and low performance when they are selected few times (in the worst case, once). Let $v \in \mathbb{V}$ be the set of template vertex indices that have at least one annotation transfer ($t_v \geq 1$) and denote the vertex set cardinality (for vertices with non-zero t_v) be $|\mathbb{V}|$. We define

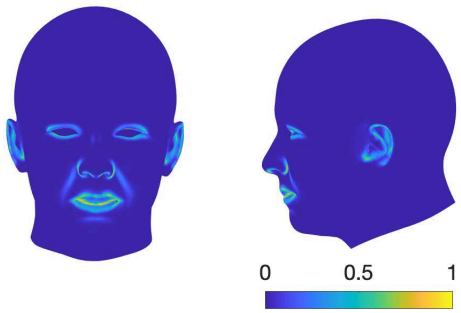


Fig. 5: Annotation transfer density color map for $N_{subj} = 675$ Headspace subjects [9] after L-ICP registration. (We omit 184 annotated subjects where there is not a full set of ear landmarks due to hair/cap occlusions.) The transfer is highly-repeatable around the mouth and nose region but is lower around the nasolabial folds.

the mean *annotation transfer density* as:

$$\bar{d} = \frac{1}{N_{subj}|\mathbb{V}|} \sum_{v \in \mathbb{V}} t_v \tag{10}$$

where N_{subj} is the number of subject target scans employed in the evaluation.

The density metric is straightforward to apply and is a quantitative measure that relates directly to qualitative annotation density colormaps. However, it is annotation-label agnostic and does not handle the case when annotation contours are in close proximity to each other (*e.g.* bottom of upper lip and top of lower lip). Here, contours with different semantic labels may transfer to the same morphed template vertices. Ideally, template vertices are selected by annotations of a single semantic label. Therefore, we additionally define a mean *annotation transfer homogeneity* metric. To do this, we define $t_{v,i} \geq 1$ as the non-zero number of annotation transfers for vertex $v \in \mathbb{V}_i$ where i indexes a semantic annotation label in the full set of annotations $\mathbb{A}$. The mean homogeneity is then

$$\bar{h} = \sum_{i \in \mathbb{A}} \omega_i \bar{h}_i = \frac{\omega_i \sum_{v \in \mathbb{V}_i} t_{v,i}}{\sum_{v \in \mathbb{V}_i} \sum_{j \in \mathbb{A}} t_{v,j}}, \quad w_i = \frac{\sum_{v \in \mathbb{V}_i} t_{v,i}}{\sum_{v \in \mathbb{V}} \sum_{j \in \mathbb{A}} t_{v,j}} \tag{11}$$

where $\bar{h}_i$ is mean homogeneity per annotation label i and ω_i is a weighting based on the relative prevalence of that annotation label, with $\sum_{i \in \mathbb{A}} \omega_i = 1$.

A limitation of these metrics is that they are template mesh specific. Templates of higher resolutions will give lower values for these metrics, since a given template vertex can only be selected for the annotation transfer (*i.e.* as the 1-nearest neighbour) over a smaller surface area. It may be possible to use an additional normalising factor, α_{res}, that adjusts for template resolution, but this may be confounded by non-uniform template resolutions.

In Figure 6 we compare L-ICP with a version of our framework that has the per-vertex affine constraint (PVAC) [1] substituted for Laplacian regularisation, with all parameters the same. We compare the two approaches in terms

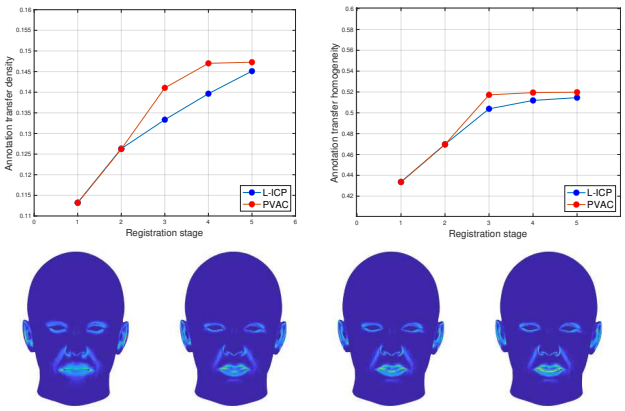


Fig. 6: Top left: annotation transfer density, 118 subjects over five registration stages. PVAC employs the *Per Vertex Affine Constraint* [1] within our framework. Top right: annotation transfer homogeneity. Bottom row: colormaps of annotation transfer density: L-ICP stages two (left) to five (right) - please zoom.

of annotation density and homogeneity over five stages of morphing. (This is done over 118 subjects due to high PVAC computation time.) Although the PVAC constraint approaches its maximum in fewer stages, the final performance is very similar at a fraction of the computational cost.

C. Processing time

The following table gives the average processing times in seconds for L-ICP to reach the *end* of each stage (averaged over 118 scans). Scans are typically 150K-200K vertices with a template of size 11.51K vertices. This was evaluated on a Macbook Pro with 2.3 GHz Quad-Core Intel Core i7, 32GB of memory, macOS Big Sur, running Matlab version R2021a. L-ICP is over 26 times faster at stage 5, than when a per-vertex affine constraint (PVAC) is employed within the same coarse-to-fine framework, with the same features and using the same stiffness schedule. Thus, L-ICP gives a dense, consumer laptop-based registration in around 47 seconds compared to around 20 minutes for PVAC.

Stage	1	2	3	4	5
L-ICP (s)	0.03	0.28	15.68	32.38	47.13
PVAC (s)	0.03	0.24	252.86	1183.67	1230.51

VII. Conclusions

We have demonstrated that fully-automatic, *densely-corresponded* non-rigid registration only requires a Laplacian regularisation term and hence is relatively rapid to compute. It achieves this via a *small deformation per iteration* assumption within a progressive coarse-to-fine framework that is guided by within-set correspondences from application-specific feature extractors. We have shown that the results are comparable with using per-vertex affine constraints, but at considerably lower computational cost. Finally, we presented a new benchmark for registration based on contour sketch annotations and a pair of annotation transfer metrics.

References

[1] Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid ICP algorithms for surface registration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition (2007)

[2] Benz, M., Laboureux, X., Maier, T., Nkenke, E., Seeger, S., Neukam, F., Häusler, G.: The symmetry of faces. In: Proceedings of Vision, Modeling and Visualization. pp. 43–50. VMV '02 (2002)

[3] Besl, P., McKay, N.D.: A method for registration of 3D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992)

[4] Bogo, F., Romero, J., Loper, M., Black, M.J.: FAUST: Dataset and evaluation for 3D mesh registration. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, Piscataway, NJ, USA (Jun 2014)

[5] Bolkart, T., Wuhrer, S.: A Groupwise Multilinear Correspondence Optimization for 3D Faces. In: IEEE International Conference on Computer Vision (ICCV). pp. 3604–3612. IEEE (2015)

[6] Bookstein, F.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(6), 567–585 (1989)

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

[8] Dai, H., Pears, N., Smith, W.: Non-rigid 3d shape registration using an adaptive template (Sep 2018), eCCV 2018: PeopleCap Workshop ; Conference date: 14-09-2018 Through 14-09-2018

[9] Dai, H., Pears, N., Smith, W., Duncan, C.: Statistical modeling of craniofacial shape and texture. International Journal of Computer Vision **128**(2), 547–571 (2019)

[10] Dai, H., Pears, N., Smith, W.: A data-augmented 3D morphable model of the ear. In: The 13th IEEE Int. Conf. on Automatic Face and Gesture Recognition. FG 2018 (2018)

[11] Erler, P., Guerrero, P., Ohrhallinger, S., Mitra, N.J., Wimmer, M.: Points2Surf: Learning implicit surfaces from point clouds. In: Computer Vision – ECCV 2020. pp. 108–124 (2020)

[12] Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: 3D-CODED: 3D correspondences by deep deformation. In: Computer Vision – ECCV 2018. pp. 235–251 (2018)

[13] Huang, X., Yang, H., Vouga, E., Huang, Q.: Dense correspondences between human bodies via learning transformation synchronization on graphs. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 17489–17501. Curran Associates, Inc. (2020)

[14] Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition pp. 1867–1874 (2014)

[15] Koo, B., Özgür, E., Le Roy, B., Buc, E., Bartoli, A.: Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) Medical Image Computing and Computer Assisted Intervention, MICCAI 2017. pp. 326–334 (2017)

[16] Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. In: Proceedings of the Symposium on Geometry Processing. p. 1421–1430. SGP '08, Eurographics Association (2008)

[17] Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) **36**(6), 194:1–194:17 (2017)

[18] Li, T., Liu, S., Bolkart, T., Liu, J., Li, H., Zhao, Y.: Topologically consistent multi-view face inference using volumetric sampling. In: Proc. International Conference on Computer Vision (ICCV). pp. 3824–3834 (Oct 2021)

[19] Liu, F., Liu, X.: Learning implicit functions for topology-varying dense 3D shape correspondence. In: In Proceeding of 2020 Conference on Neural Information Processing Systems (2020)

[20] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia) **34**(6), 248:1–248:16 (2015)

[21] Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(12), 2262–2275 (2010)

[22] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning continuous signed distance functions for shape representation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

[23] Ploumpis, S., Ververas, E., O'Sullivan, E., Moschoglou, S., Wang, H., Pears, N., Smith, W., Gecer, B., Zafeiriou, S.P.: Towards a complete 3D morphable model of the human head. IEEE Trans. Pattern Analysis and Machine Intelligence (2020)

[24] Sandro, L., Bangbang, Y., Tianxing, F., Hujun, B., Guofeng, Z., Marc, P., Zhaopeng, C.: Latenthuman: Shape-and-pose disentangled latent representation for human bodies (2021)

[25] Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C., Seidel, H.P.: Laplacian surface editing. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. p. 175–184. SGP '04, Association for Computing Machinery (2004)

[26] Styner, M.A., Rajamani, K.T., Nolte, L.P., Zsemlye, G., Székely, G., Taylor, C.J., Davies, R.H.: Evaluation of 3D correspondence methods for model building. In: Biennial Int. Conf. on Information Processing in Medical Imaging. pp. 63–75 (2003)

[27] Sun, H., Pears, N., Dai, H.: A human ear reconstruction autoencoder. In: Proc. 16th Int. Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. pp. 136–145. VISAPP '20 (2020)

[28] Sung, M., Su, H., Yu, R., Guibas, L.: Deep functional dictionaries: Learning consistent semantic structures on 3d models from functions. In: NeurIPS (2018)

[29] Yang, J.: The thin plate spline robust point matching (TPS-RPM) algorithm: A revisit. Pattern Recogn. Lett. **32**(7), 910–918 (2011)

[30] Zheng, Z., Yu, T., Dai, Q., Liu, Y.: Deep implicit templates for 3D shape representation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1429–1439 (2021)