Eurographics Workshop on 3D Object Retrieval (2019)
SHREC - 3D Shape Retrieval Contest (2019)
S. Biasotti, G. Lavoué, B. Falcidieno, and I. Pratikakis (Editors)
R. C. Veltkamp (SHREC organizer)

SHREC 2019: Matching Humans with Different Connectivity

S. Melzi[1], R. Marin[1], E. Rodolà[2], U. Castellani[1], J. Ren[3], A. Poulenard[4], P. Wonka[3], M. Ovsjanikov[4]

[1]University of Verona [2]Sapienza University of Rome [3]KAUST University [4]LIX, Ecole Polytechnique

Abstract

Objects Matching is a ubiquitous problem in computer science with particular relevance for many applications; property transfer between 3D models and statistical study for learning are just some remarkable examples. The research community spent a lot of effort to address this problem, and a large and increased set of innovative methods has been proposed for its solution. In order to provide a fair comparison among these methods, different benchmarks have been proposed. However, all these benchmarks are domain specific, e.g., real scans coming from the same acquisition pipeline, or synthetic watertight meshes with the same triangulation. To the best of our knowledge, no cross-dataset comparisons have been proposed to date. This track provides the first matching evaluation in terms of large connectivity changes between models that come from totally different modeling methods. We provide a dataset of 44 shapes with dense correspondence as obtained by a highly accurate shape registration method (FARM). Our evaluation proves that connectivity changes lead to Objects Matching difficulties and we hope this will promote further research in matching shapes with wildly different connectivity.

CCS Concepts

• ***Computing methodologies*** → ***Shape analysis;*** • ***Theory of computation*** → *Computational geometry;*

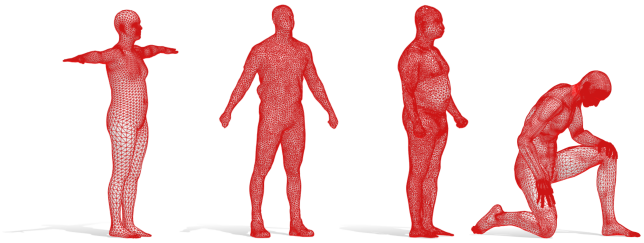


Figure 1: *Sample meshes involved in this track. The shapes differ in terms of mesh resolution as well as for triangles and vertex distribution. Several additional variations are also considered.*

1. Introduction

Recent technological advances provide new modeling techniques, enlarging the set of applications and involving a wider mass of consumers [PLB12]. Modeling software enables artists to easily deform shapes, perform surface remeshing and make models ready for real-time animation [Hor18]. Moreover, off-the-shelf sensing devices put 3D body scanning technology at disposal of everyone [ZSG*18]. These facts have led to a wide production of 3D models with different resolution, sampling density, distribution of details, noise artifacts, and so forth [BRLB14, LMB14, BBK08, CGF09].

With this track we evaluate different matching pipelines and descriptors over a collection of shapes originating from a diverse set of datasets. This entails dealing with different surface discretizations, as well as other types of nuisance such as the presence of disconnected components, noisy or weakly cluttered surfaces (e.g., due to clothes and accessories), and a wide range of topology and geometry dissimilarities. The evaluated methods are compared in terms of correspondence coherence, runtime, and other implementation details. Only one group participates to our track, and we propose two alternatives to the proposed matching pipeline. Furthermore, we test the stability and properties of different point descriptors in challenging setups, in an attempt to identify the most suitable for real-world applications. From our results, we observe that shape correspondence remains an open and challenging problem whenever connectivity changes are present in the data, and conclude that a concerted effort is required in this direction. This report is accompanied by dense ground-truth correspondences and evaluation code to foster further research[†].

2. Digital humans representation

Human body is among the most common subjects of 3D digital acquisition, arising in a wide range of applications. Arguably the most common digital representation for human shapes is as a polygonal mesh, where one has to face the choice of *mesh connectivity*, giving rise to extreme variations in the representation.

We model human body as a surface $\mathcal{M} \subset \mathbb{R}^3$, possibly with a boundary $\partial\mathcal{M}$ in case of missing parts, self-occlusions or other imperfections in the acquisition process. We represent $\mathcal{M}$ as a *manifold triangle mesh* in the discrete setting, composed by set of n vertices $\mathcal{V} \in \mathbb{R}^{n\times 3}$ connected by edges $\mathcal{E}$; each edge belongs to at most two triangles. The collection of these triangles approximates the underlying surface, and encodes the mesh *connectivity*.

[†] http://profs.scienze.univr.it/~marin/shrec19/

DOI: 10.2312/3dor.20191070

We further equip the surface $\mathcal{M}$ with the *Laplace-Beltrami Operator* (LBO), a second-order differential operator generalizing the classical Laplacian to manifolds. In the discrete setting, it admits a representation as a $n \times n$ matrix $\boldsymbol{\Delta} = \boldsymbol{A}^{-1}\boldsymbol{W}$, where $\boldsymbol{A}$ is the *mass matrix* and $\boldsymbol{W}$ is the *stiffness matrix* defined accordingly to [PP93]. The LBO is a positive semi-definite operator $\boldsymbol{\Delta}_{\mathcal{M}} \colon \mathcal{F}(\mathcal{M},\mathbb{R}) \to \mathcal{F}(\mathcal{M},\mathbb{R})$ admitting an eigendecomposition $\boldsymbol{\Delta}_{\mathcal{M}}\boldsymbol{\phi}_l = \lambda_l \boldsymbol{\phi}_l$, with real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots$ and corresponding eigenfunctions $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots$, corresponding to the Fourier basis on manifolds.

3. The proposed benchmark

The dataset we provide consists of 430 shape pairs. The shapes themselves come from different sources, and dense *cross-dataset* correspondences are obtained with the registration method FARM [MMRC18]. We remark that correspondence between human bodies is ill-posed and a unanimous definition is hard to achieve; indeed, even when experts are involved, usually only sparse key points are provided. By using FARM, our dense correspondences adhere to the semantics of the SMPL human template [LMR*15].

3.1. Source datasets

SMPL [LMR*15] is a vertex-based 3D morphable model, learned statistically over a wide population of real scan data. The human shape is controlled by 82 parameters (of which 10 for identity and 72 for pose). The SMPL model provides a simple way to generate realistic bodies that share the same connectivity.

FAUST [BRLB14] is a benchmark of real scans of different humans, coming with ground-truth correspondence for a subset. Each shape is obtained by annotating a human subject by anthropometry experts, using 17 bones as sparse key points. The SMPL template is aligned to the scans, providing a dense map. Both the aligned template and the real scans are available. This dataset provides interesting challenges, including acquisition noise, holes, and self-contact.

SCAPE [ASK*05] is a milestone in human body registration. This pipeline fits a template to real data by solving a complex optimization problem over triangle faces. The method is able to capture different body shapes, and is capable to work with different data representations (e.g., range maps and mocap markers).

TOSCA [BBK08] high-resolution is a synthetic dataset with non-rigid deformations of shapes from different classes. All shapes have around $50K$ vertices, and models of the same class are in correspondence. TOSCA shapes are a good example of handcrafted objects.

SPRING is a dataset generated by modifying a template using parameters carrying anthropometric semantics (e.g., height, calf circumference, etc.). The body shape space is learned by registering a template in SCAPE fashion to over $3K$ body models. Then, PCA is used to find directions with a clear and useful meaning. The dataset consits of a high variety of shapes in full correspondence.

MoSh Mocap [LMB14] is a dataset produced from motion capture data acquisition, with soft tissue information yielding comparable quality with full-fledged 3D body scanners. The dataset provides clean real data with highly regular tessellation, reflecting nowadays' expectations of real acquisitions.

BadKing.com.au is a website collecting contributions from professional artists which are made freely available. To our knowledge, correspondence methods have never been compared with this sort of data. The meshes present high levels of detail, disconnected components, holes, and an enormous amount of different styles.

CAESAR [RDP99] is a rich real body scans set. It has been used widely in data-driven works, and is nowadays the baseline to learn generative human body models. Unfortunately, it is not freely available and redistribution is limited. For this reason we rely over [PWH*17] that provides registrations of this dataset to the research community. All shapes are in correspondence and in a neutral pose.

Princeton [CGF09] is a segmentation benchmark built on top of the SHREC 2007 Watertight Models track. Its shapes come from different sources and include synthetic human bodies with robot-like proportions as well as noisy real scans. They span both low ($\sim 4.7K$ vertices) and moderately high resolutions ($15K$ and more).

SHREC14 [PSR*16] Shape Retrieval of Non-Rigid 3D Human Models track has two subsets. A realistic one, with CAESAR shapes registered using SCAPE and remeshed to $\sim$15K vertices; and a synthetic one with plastic poses, a smooth surface, many details and hand articulations ($\sim$60K) vertices.

K3D-Hub [XZC18] provides a method to register a high-quality template to a low-quality Kinect scan. The low-resolution setup provides some interesting challenges: subjects may be clothed, with few details but with dense ($\sim 10K$ vertices) and regular connectivity. Matching these low-res shapes to more detailed ones may have interesting applications in entertainment.

3.2. Connectivity variations

Every data source comes with unique characteristics: different purposes require different modelling principles, in turn affecting connectivity. With this challenge, we want to encourage the community to take this type of tricky variations into consideration.

Different orientation. To avoid relying upon a-priori knowledge on the position in ambient space, the shapes in our composite dataset are *not* pre-aligned into a coherent orientation.

Connectivity artifacts. There are shapes with broken or missing connectivity (e.g., outlier points belonging to no triangle). We propose to take into consideration these scenarios.

Different density. All these shapes have different resolution and triangle density. This is particularly challenging for methods that rely upon similar discretization. Artists create models with a clean and optimized meshing, with few degenerate triangles and different density depending on the surface region. Real scans on the other hand may result from a complex surface reconstruction pipeline giving rise to degenerate triangles and non-manifold artifacts. Fitted templates either assume uniform density (e.g., SCAPE triangles are equally distributed over the surface), or provide more detail around salient points (e.g., SMPL is denser on the human face).

Additional variations. We also consider variations of **identity** and **pose**, and include **different surface artifacts** such as **topological noise**, clothes, hair or accessories. We analyze both watertight meshes and meshes with disconnect components.

4. Methods

Point-to-point matching is performed i) by computing point descriptors and matching those in descriptor space, or ii) by designing a matching pipeline. We evaluated 7 descriptors and 3 pipelines.

4.1. Descriptors

Point descriptors characterize the neighborhood of each point on a discrete surface, and are expected to be (1) discriminative (different points should have different descriptors); (2) repeatable under noise and deformation; (3) fast to compute; and (4) compact.

Given two descriptor fields $DESC_{\mathcal{M}}$ and $DESC_{\mathcal{N}}$ on shapes $\mathcal{M}$ and $\mathcal{N}$ respectively, a point correspondence for each $x \in \mathcal{M}$ can be obtained by a nearest-neighbor search in descriptor space:

$$y^* = \arg\min_{y \in \mathcal{N}} \|DESC_{\mathcal{M}}(x) - DESC_{\mathcal{N}}(y)\|_F \,. \tag{1}$$

We use the equation above in our tests. To eliminate differences due to scale, all meshes are rescaled to similar surface area.

GPS. The *Global Point Signature* (GPS) [Rus07] is a point descriptor defined as the q-dimensional vector:

$$GPS(x) = [\lambda_2^{-\frac{1}{2}} \phi_2(x), \ldots, \lambda_{q+1}^{-\frac{1}{2}} \phi_{q+1}(x)] \,, \tag{2}$$

in our experiments, we set $q = 100$.

HKS. The *Heat Kernel Signature* (HKS) [SOG09] is built upon the heat kernel between a point and itself, expressed in a (truncated) spectral decomposition as $k_t(x,x) = \sum_{l=1}^{K} e^{-t\lambda_l} \phi_l(x)^2$. This can be interpreted as the amount of heat that remains at point x after a delta distribution is diffused for time t. Given a fixed set of time scales $\{t_1, \ldots, t_q\} \in \mathbb{R}^q$, the HKS at a point x is defined as:

$$HKS(x) = [k_{t_1}(x,x), \ldots, k_{t_q}(x,x)]. \tag{3}$$

We consider $q = 100$ as suggested in [SOG09] and $K = 200$.

WKS. The *Wave Kernel Signature* (WKS) [ASC11] extends the ideas above by modeling a quantum particle on the surface with a given initial energy E. The descriptor for point $x \in \mathcal{M}$ is defined as the average probability over time to find the particle at position x, and is computed as $wks_E(x) = \sum_{l=1}^{K} f_E(\lambda_l)^2 \phi_l(x)^2$, where $f_E(\lambda_l)^2$ is log-normal energy probability distribution. Given a set of energy levels $\{E_1, \ldots, E_q\}$, the WKS is defined as:

$$WKS(x) = [wks_{E_1}(x), \ldots, wks_{E_q}(x)] \,. \tag{4}$$

We use $K = 200$ basis functions and $q = 100$ energy levels.

AWFT [MRCB16] is based on the definition of Anisotropic Windowed Fourier transform on non-Euclidean domains, and uses the Anisotropic LBO [ARAC15]. A family of such operators is defined depending on two parameters, anisotropy α and orientation θ. A Gaussian window $g^{\tau}_{x,\alpha,\theta}$ is expressed for given α, θ in the Anisotropic LBO basis, with variance $\tau > 0$, translated to each vertex and modulated with respect to the K smallest Laplacian eigenvalues. Given a scalar function $f : \mathcal{M} \to \mathbb{R}$, the coefficients of its windowed Fourier transform $(Sf)^{\tau}_{x,l,\alpha,\theta}$ are given by the inner product between f and the atoms $g^{\tau}_{x,\lambda_l,\alpha,\theta}$. Dependence on parameter l is removed in [MRCB16] via application of the total weighted power, aggregating in a single value $(S_{TWP}f)^{\tau}_x$ all coefficients with different modulation. For a given point $x \in \mathcal{M}$, its AWFT is:

$$AWFT(x) = [(S_{TWP}f)^{\tau_1}_{x,\alpha_1,\theta_1}, \ldots, (S_{TWP}f)^{\tau_T}_{x,\alpha_A,\theta_O}] \,. \tag{5}$$

We use $K = 200$ eigenfunctions and the parameters τ, α, θ are fixed as suggested in [MRCB16], obtaining a 100-dimensional descriptor. For increased efficiency, we remesh via edge collapse [GH97] all shapes with $> 60K$ vertices and then extend the matches to full resolution via nearest-neighbors in $\mathbb{R}^3$.

DEP. The *discrete-time evolution process* (DEP) [MOR*18] encodes the action of an integral operator on the surface. This action is defined on top of a pairwise potential $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+$ that depends on the geometry of the surface and encodes the degree of influence that surface points exert on each other:

$$Af_{(t)} = \int_{\mathcal{M}} d(\cdot, y) f_{(t)}(y) dy \,, \tag{6}$$

for scalar functions $f_{(t)} : \mathcal{M} \to \mathbb{R}$. The action of d is realized by the following recursive relations:

$$f_{(t+1)} = A f_{(t)} \,. \tag{7}$$

A score is defined for a fixed number T of time steps as $s(x) = f_0(x) + \sum_{t=1}^{T} A^t f_0(x)$, summing up the contributions of the evolution process (7) across all discrete times $t = 1, \ldots, T$. A DEP descriptor is obtained by letting $T \to \infty$ and using a multiscale approach on the choice of the pairwise potential d, giving $DEP(x) = [S_{d_1}(x), \ldots, S_{d_K}(x)]$. In our experiments we use the biharmonic distance [LRF10] for the definition of d, approximated with $K = 200$ eigenfunctions and the parameters of [MOR*18], resulting in a 100-dimensional descriptor. For meshes with $> 7K$ vertices, we only consider $7K$ farthest point samples and map the matches back to full resolution through nearest-neighbors in Euclidean space.

SHOT. The *SHOT* descriptor [TSDS10] encodes histograms of normals, which are more representative of the local structure of the surface than plain 3D coordinates. This descriptor is built on top of a stable *Local Reference Frame* (LRF) defined as the principal eigenvector of a modified covariance matrix around each point. An isotropic spherical grid with 32 partitions is aligned to the computed LRF, and the 3D distribution of the normals is represented as a local histogram per partition. The ordered concatenation of these histograms defines the descriptor at each point. We use the standard parameters of [TSDS10], yielding 320-dimensional descriptors.

GFrames SHOT is a variant of SHOT constructed on top of a novel, more stable LRF as proposed in [MST*19]. The GFrames LRF is based on the computation of the gradient of a scalar function defined on the surface. By varying the scalar function, it is possible to produce several LRFs depending on the desired stability properties. Following [MST*19], we adopt the square of the first non-constant Laplacian eigenfunction as a scalar function. We refer to the resulting 320-dimensional descriptor as GSHOT.

4.2. Matching pipelines

Functional Maps [OBCS*12, OCB*17] are based on the idea that seeking for functional (as opposed to point-to-point) correspondences makes the problem independent on the shape discretization

and easier to optimize. Any point-to-point map $\mathbf{\Pi}: \mathcal{N} \to \mathcal{M}$ induces a functional correspondence $\boldsymbol{T}: \mathcal{F}(\mathcal{M},\mathbb{R}) \to \mathcal{F}(\mathcal{N},\mathbb{R})$ via the composition $\boldsymbol{T}(\boldsymbol{f}) = \boldsymbol{f} \circ \mathbf{\Pi}$. In the Laplacian eigenbases $\mathbf{\Phi}, \mathbf{\Psi}$, $\boldsymbol{T}$ is represented as a matrix $\boldsymbol{C} \in \mathbb{R}^{K_{\mathcal{N}} \times K_{\mathcal{M}}}$ that maps coefficients w.r.t. the basis $\mathbf{\Phi}$ to coefficients w.r.t. $\mathbf{\Psi}$. Following [NO17], we estimate a functional map $\mathbf{C}$ by solving the non-convex problem:

$$\arg\min_{\mathbf{C}} \|\mathbf{C}\hat{\mathbf{F}} - \hat{\mathbf{G}}\mathbf{C}\|_F^2 + w_1\|\mathbf{C}\hat{\mathbf{F}} - \hat{\mathbf{G}}\|_F^2 + w_2\|\mathbf{C}\mathbf{\Lambda}_{\mathcal{M}} - \mathbf{\Lambda}_{\mathcal{N}}\mathbf{C}\|_F^2 \tag{8}$$

where $\mathbf{\Lambda}_{\mathcal{M}}, \mathbf{\Lambda}_{\mathcal{N}}$ are diagonal matrices of the Laplacian eigenvalues. Matrices $\hat{\mathbf{F}} \in \mathbb{R}^{k_{\mathcal{M}} \times q}, \hat{\mathbf{G}} \in \mathbb{R}^{k_{\mathcal{N}} \times q}$ contain the Fourier expansion coefficients of q probe functions on $\mathcal{M}$ and $\mathcal{N}$. We use 20-dimensional WKS descriptors concatenated with 20-dimensional wave kernel maps [OBCS*12] around body landmarks detected as in [MMRC18]. We set $k_{\mathcal{M}} = 60$ and $k_{\mathcal{N}} = 60$. Given a functional map matrix $\mathbf{C}$, the underlying point-wise map $\mathbf{\Pi} \in \{0,1\}^{n_{\mathcal{N}} \times n_{\mathcal{M}}}$ is recovered by solving the projection problem [OCB*17]:

$$\min_{\mathbf{\Pi}} \|\mathbf{C}\mathbf{\Phi}^\top - \mathbf{\Psi}^\top\mathbf{\Pi}\|_F^2 \quad \text{s.t. } \mathbf{\Pi}^\top\mathbf{1} = \mathbf{1}. \tag{9}$$

Iteratively Refined Functional Maps (bFMAP) follows the map refinement and estimation method of [LRR*17, MMRC18]. Given an input functional map $\boldsymbol{C}^{(0)} = \boldsymbol{C}$, mismatches are filtered out by solving a sequence of *convex* problems:

$$\mathbf{C}^{(t+1)} = \arg\min_{\mathbf{C}} \|\mathbf{C}^{(t)}\hat{\mathbf{F}}^{(t)} - \hat{\mathbf{G}}^{(t)}\|_{2,1} + \mu\|\mathbf{C}^{(t)} \circ \mathbf{W}\|_F^2, \tag{10}$$

with $t = 0, \dots, T$. The μ-term enforces a diagonal structure on $\mathbf{C}$, where the shape of the diagonal is encoded in the "mask" matrix $\mathbf{W}$. As probe functions we use pairs of corresponding deltas $(\delta_{x_i}^{\mathcal{M}}(x), \delta_{\pi^{(0)}(x_i)}^{\mathcal{N}}(y))_{i=1}^{q}$, where $\pi^{(0)}$ is the point-wise conversion of $\boldsymbol{C}^{(0)}$ via (9). The $\ell_{2,1}$ norm promotes column-wise sparsity, allowing to downweigh mismatches during the refinement process.

As done in [MMRC18], we set $\mu = 0.01$, $T = 20$ iterations, and $q = 1000$ uniformly distributed delta functions over $\mathcal{M}$.

BCICP is the only method submitted to our track from the authors of [RPWO18]. It is a recent algorithm for functional map estimation, employing an *orientation-preserving* regularizer and a new refinement procedure named Bijective and Continuous ICP (BCICP).

Orientation-preserving regularizer. Given two shapes $\mathcal{M}$ and $\mathcal{N}$, and a set of q pairs of probe functions $\{(f_i, g_i)\}_{i=1}^{q}$, a data term is setup as in (8). Then, the following regularizer is introduced:

$$E_{\text{orient}} = \sum_{i=1}^{k} \left\|\boldsymbol{C} \circ \Omega_{f_i} - \Omega_{g_i} \circ \boldsymbol{C}\right\|_F^2, \tag{11}$$

where Ω is an operator that extracts the *orientation* of a local frame at each point, as encoded by the surface normal and the gradients of the given descriptors. Eq. (11) attempts to preserve the orientation of every corresponding local frame induced by the descriptors.

BCICP refinement. Similarly to ICP, the refinement alternatively solves for a point-wise map and a functional map. However, this happens *both* in the spectral domain and the spatial domain by making use of several heuristics as follows:

Continuity of the point-wise map is improved by smoothing out the displacement vector field induced by the map and filtering out the outlier regions. Assume $\pi: \mathbf{x}_i \mapsto \mathbf{y}_{\pi(i)}$ from source to target shape. To smoothen the correspondence at a vertex, we smooth out the associated displacement vector $\mathbf{t}_i = \mathbf{y}_{\pi(i)} - \mathbf{x}_i$ using the neighboring ones. Edges are classified as 'outliers' if the mapped endpoints have large distance, since they are likely to be the boundary of outlier regions; such edges are removed from the mesh adjacency matrix, and points that do not belong to the largest connected component of the modified connectivity will be regarded as outliers.

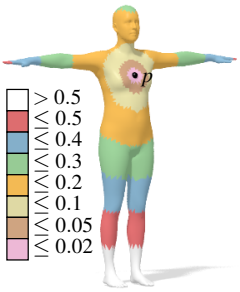
Bijectivity is improved by considering extra energies defined on the compound point-wise maps from both sides. Specifically, the original ICP uses the energy $E(C_{\mathcal{MN}}, \pi_{\mathcal{NM}}) = \left\|\mathbf{\Psi}C_{\mathcal{MN}} - \pi_{\mathcal{NM}}\mathbf{\Phi}\right\|^2$, where $C_{\mathcal{MN}}$ is the functional map from $\mathcal{M}$ to $\mathcal{N}$, and $\pi_{\mathcal{NM}}$ is the associated point-wise map from $\mathcal{N}$ to $\mathcal{M}$. To promote bijectivity, the modified energy $\hat{E}(C_{\mathcal{MM}}, \pi_{\mathcal{MN}}, \pi_{\mathcal{NM}}) = \left\|\Phi_1 C_{\mathcal{MM}} - \pi_{\mathcal{MN}}\pi_{\mathcal{NM}}\mathbf{\Phi}\right\|_F^2$ is used, where the auxiliary variable $C_{\mathcal{MM}}$ is a functional map from shape $\mathcal{M}$ to itself. This energy helps to regularize the compound map $\pi_{\mathcal{MN}}\pi_{\mathcal{NM}}$ to be identity. A similar term for $\pi_{\mathcal{NM}}\pi_{\mathcal{NM}}$ is also added to the total energy.

Finally, map **coverage** is improved by spreading out the correspondences of vertices with a large pre-image. A vertex on the target shape is "covered" if it is the image of at least one vertex on the source shape. More discussion can be found in [RPWO18].

For the application of BCICP, meshes are downsampled to $\approx 5K$ vertices using [YBZW14]. BCICP is then executed with the default parameters and 10 iterations per pair. The estimated maps are propagated back to the original shapes by simple nearest-neighbor search; therefore, the final maps may have low coverage.

5. Evaluation

We adopt the standard error measure defined in [KLF11]. Each method is represented as a curve denoting the percentage of correspondences (y-axis) with (normalized) geodesic error below a varying threshold (x-axis). We only plot the interval $[0, 0.5]$, while the *average geodesic error* (AGE) considers the interval $[0, 1]$. The normalized distances from a point p are shown in the inset figure.



Ground truth. The ground-truth is given by the state-of-the-art registration method FARM [MMRC18]. We use it to register all the shapes to SMPL, obtaining as a side-product a meaningful dense map in both directions. A map between two given shapes $\mathcal{M}$ and $\mathcal{N}$ is then obtained by composing the map from $\mathcal{M}$ to SMPL with the one from SMPL to $\mathcal{N}$ (see two examples in Figure 2).

Approximate geodesic error. Since our dataset includes several meshes with hundreds of thousands of vertices, we approximate geodesic error. Given two shapes $\mathcal{M}, \mathcal{N}$ and a correspondence:

- We only consider the subsets of 6890 vertices of the SMPL model registered via FARM, denoted by $\mathcal{M}_v$ and $\mathcal{N}_v$.
- We denote by $\widetilde{\mathcal{N}}_v$ the estimated matches for the points in $\mathcal{M}_v$.

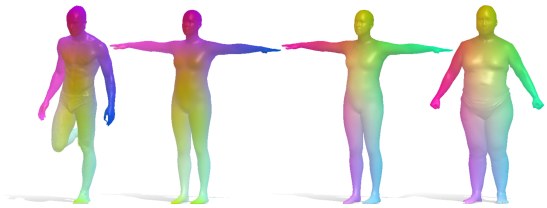


Figure 2: *Two examples of dense point-to-point correspondence provided by the FARM registration pipeline. The model in T-pose is the SMPL template. Correspondence is encoded by colors.*

- We compute distances between the 6890 vertices of $\mathcal{N}_v$ and the 6890 vertices of $\widetilde{\mathcal{N}}_v$ using Dijkstra's algorithm.
- For points located on disconnected components (e.g. due to partiality or accessories) we use Euclidean distances.
- Distances are normalized to within $[0,1]$.

6. Results

The dataset of 430 pairs is partitioned into separate (and possibly overlapping) subsets described in the following.

Others vs SMPL. We test the capability to map different human bodies to a common template. Here we measure the stability to noise over the *source* shape (43 pairs).

SMPL vs Others. This measures how noise over the *target* shape affects the method's performance (43 pairs).

with SMPL. The data here is a combination of pairs from the two previous sets. This simulates a more realistic setting in which source and target do not have any a-priori role (86 pairs).

Others vs Others. The SMPL template never appears, hence one cannot rely on any expectation of mesh regularity (344 pairs).

Different Connectivity. This is the core of our challenge. Differently from the *Others vs Others* experiment, here we do not allow pairs from the same dataset. Therefore, *all* shape pairs have different connectivity (415 pairs).

Different Connectivity plus Symmetry. Same as above, but for each point we consider correct both the ground-truth correspondence and its symmetric counterpart (415 pairs).

Same Connectivity. With this reduced set, we evaluate how much methods improve by exploiting this assumption (15 pairs).

Same Connectivity plus Symmetry. same pairs evaluated for the Same Connectivity case, but considering also the symmetric points of the ground-truth correspondence as correct (15 pairs).

All pairs. The complete dataset, unifying all previous experiments. In Figure 5, we povide a visual summary of this set (430 pairs).

6.1. Comparisons

Descriptors. In Figure 3 we compare the descriptors of Section 4.1 in all the settings detailed above. The worst results overall are obtained by GPS, while WKS is consistently the best except for the *Same Connectivity* case. HKS is second best, followed by AWFT.

Descriptor	Avg AGE	Max AGE	Max AGE Pair
AWFT	0.33	0.50	27_40
DEP100	0.38	0.73	10_16
GPS100	0.35	0.63	22_40
HKS100	0.25	0.49	37_40
SHOT	0.35	0.47	18_40
WKS100	0.21	0.54	1_40
GSHOT	0.34	0.49	39_18
Corr Methods			
FMAP	0.27	0.75	39_10
bFMAP	0.27	0.75	34_10
BCICP	0.08	0.57	12_40

Table 1: *Descriptor (top) and matching pipelines (bottom) results. The last column reports the shape pair achieving the max AGE.*

We found that the latter seems quite sensitive to the specific setting, with variations in quality even among *SMPL vs Other* and *Other vs SMPL*. **All methods perform significantly better in the case of same connectivity**. DEP seems the most sensitive overall, with a dramatic drop in accuracy according to the symmetric evaluation.

Table 1 (top) reports a summary quantitative evaluation, largely confirming the remarks above. The largest error is observed for pairs that involve mesh n. 40 (depicted in Figure 4), as also confirmed in Table 2 (showing the 5 pairs with largest AGE). In Figure 5 we further plot the complete set of curves (one per shape pair) for each method. We find this visualization informative, as the curves for the SHOT, GSHOT, HKS, and WKS descriptors exhibit less spread and are more concentrated around their mean, while for AWFT, GPS, and DEP the curves are less repeatable.

Matching pipelines. In Figure 3 we also report comparisons for the matching pipelines, where BCICP comes out as the best performing method. We attribute this gap in performance to its special regularizers, and further note that the accuracy of BCICP does not directly depend on mesh resolution, since it mainly relies on geodesic distances that are not affected too much by changes in mesh connectivity. bFMAP outperforms standard FMAP on the pairs that involve the SMPL shape, while it does not give any improvement in the other cases. This is mainly due to its parameters being optimized for the registration performed by FARM, which operates toward the SMPL template. A quantitative comparison between the three pipelines is better summarized in Table 1 (bottom). In addition, in Table 3 we show comparisons on the 5 pairs with the worst AGE. Differently from the case of descriptors, however, here we do not observe any consistently difficult shape. Finally, in Figure 5 we plot curves for all the shape pairs for each matching pipeline. BCICP exhibits significantly less variance, confirming the good quality of the correspondences. We further note how the bFMAP curves are more concentrated toward the top of the graph if compared to the standard FMAP pipeline, confirming its better behavior. In Figure 5 (bottom right) we also plot the distribution of shape pairs (430 in total, x axis) at increasing average geodesic error (AGE, y axis). The mean AGE over all methods is shown in blue while the minimum and maximum AGE are depicted respectively as green and red shaded areas. The vertical blue lines identify shape pairs with same connectivity: these are mainly located on the lower end of the

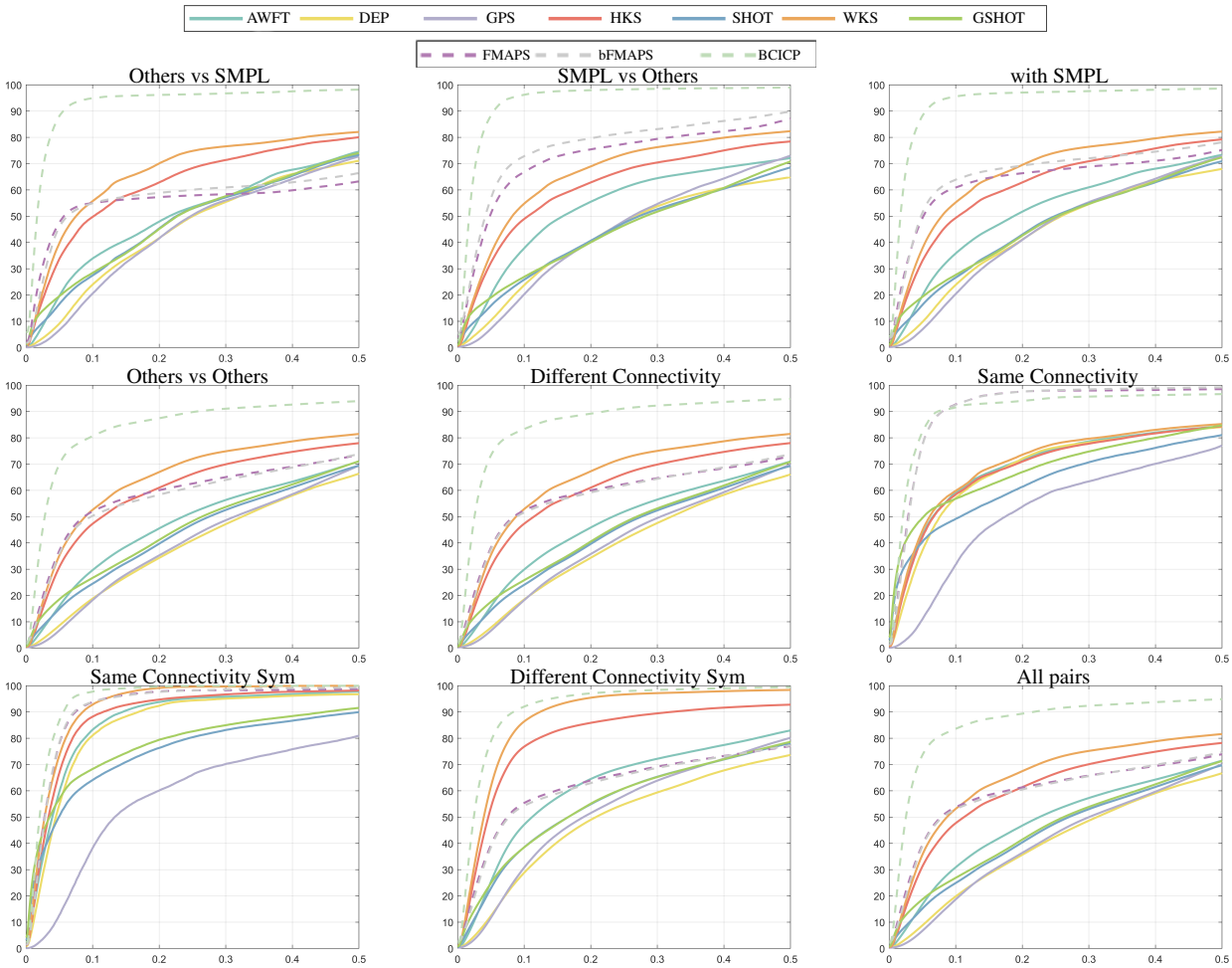


Figure 3: *Descriptor and matching pipelines comparisons. Overall, we observe that shapes with the same mesh connectivity tend to induce better correspondences.*

Pairs	Mean AGE	AWFT	DEP100	GPS100	HKS100	SHOT	WKS100	GSHOT
35_40	0.45	0.50	0.39	0.56	0.37	0.42	0.52	0.40
27_40	0.44	0.50	0.43	0.57	0.35	0.41	0.44	0.40
24_40	0.44	0.44	0.37	0.56	0.39	0.43	0.52	0.40
22_40	0.44	0.47	0.34	0.63	0.39	0.43	0.44	0.39
31_40	0.43	0.43	0.41	0.57	0.35	0.45	0.39	0.42

Table 2: *Comparison of descriptors for the 5 shape pairs with largest average geodesic error.*

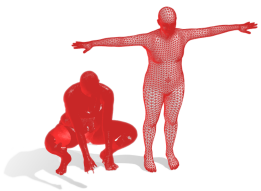


Figure 4: *A comparison between the high-resolution, non-uniform and partial mesh 40 and the SMPL template mesh.*

Pairs	Mean AGE	FMAP	bFMAP	BCICP
38_3	0.62	0.74	0.74	0.39
38_9	0.60	0.71	0.72	0.37
34_13	0.59	0.69	0.69	0.40
32_26	0.57	0.74	0.74	0.21
21_1	0.56	0.73	0.72	0.24

Table 3: *Matching pipelines on the 5 pairs with largest AGE.*

graph, meaning that estimating point-to-point correspondences for such cases is easier than for cases with different connectivity.

6.2. Runtime

All descriptors were computed on an Intel 3.6GHz i7 CPU with 32GB RAM. In Figure 6, we measure time from shape loading to descriptor storage for each shape and descriptor. FMAP has an average runtime of 28s, with a large standard deviation of 22s (worst

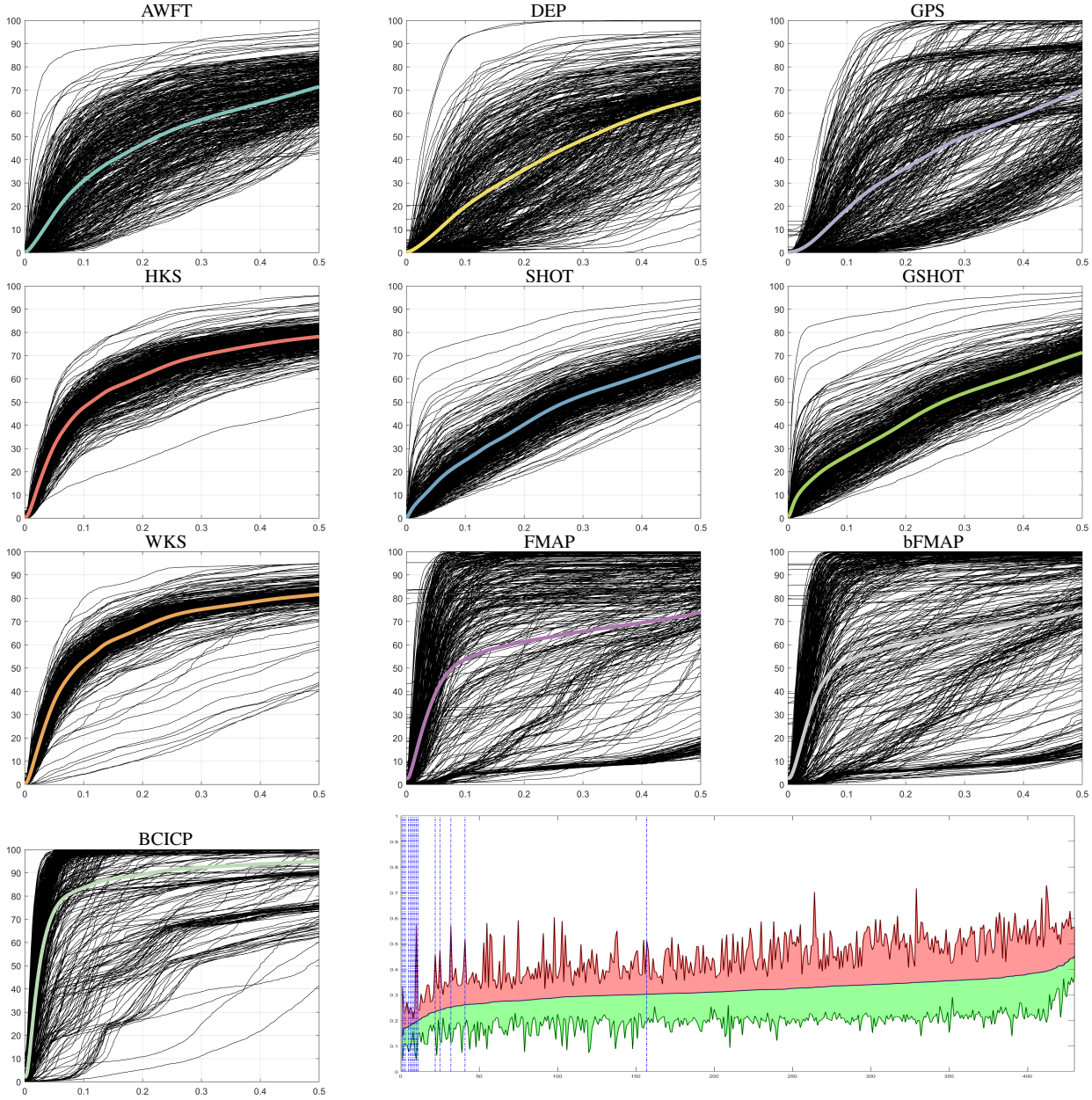


Figure 5: *Full comparisons on the entire dataset. Every black curve corresponds to a shape pair (430 per subplot); the colored curves represent the mean for each method. In the bottom right, we show the AGE (y axis) over all pairs (x axis); the blue curve is the mean AGE across all methods, while the red and green areas denote min and max AGE. The pairs with shared connectivity are identified by vertical dashed lines.*

case 151s). The average runtime for bFMAP is 427s. BCICP requires 150s of pre-processing, and 100 – 300s for map estimation.

7. Conclusion

With this SHREC track we provided a comparison among point-to-point matching algorithms for human shapes represented as triangular meshes *with different connectivity*. We demonstrate that the recent BCICP pipeline and standard descriptors such as HKS and WKS are more stable to connectivity variations which still pose a strong challenge to the shape analysis community. We conclude that, differently from common practice, past and future matching methods should be conceived and evaluated with respect to their robustness to connectivity changes. Finally, while we only considered human shapes in this track (as a consequence of using an accurate, although model-specific registration pipeline [MMRC18]),

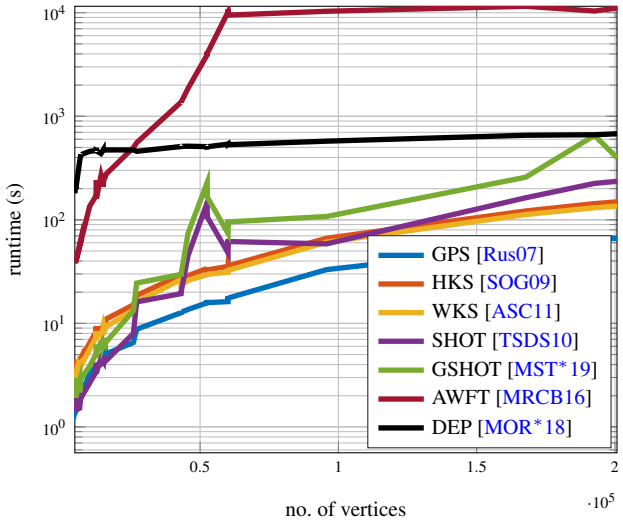


Figure 6: *Runtime comparisons for descriptor computation.*

we conjecture that our remarks on the relevance of connectivity in matching tasks may still hold for more generic shape classes.

References

[ARAC15] ANDREUX M., RODOLÀ E., AUBRY M., CREMERS D.: Anisotropic laplace-beltrami operators for shape analysis. In *Computer Vision - ECCV 2014 Workshops* (Cham, 2015), Springer International Publishing, pp. 299–312. 3

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In *Proc. ICCV Workshops* (Nov 2011), IEEE, pp. 1626–1633. 3, 8

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics 24*, 3 (July 2005), 408–416. 2

[BBK08] BRONSTEIN A., BRONSTEIN M., KIMMEL R.: *Numerical Geometry of Non-Rigid Shapes*. Springer, New York, NY, 2008. 1, 2

[BRLB14] BOGO F., ROMERO J., LOPER M., BLACK M. J.: Faust: Dataset and evaluation for 3d mesh registration. In *Proc. CVPR* (Washington, DC, USA, 2014), IEEE, pp. 3794–3801. 1, 2

[CGF09] CHEN X., GOLOVINSKIY A., FUNKHOUSER T.: A benchmark for 3D mesh segmentation. *TOG 28*, 3 (aug 2009), 73:1–73:12. 1, 2

[GH97] GARLAND M., HECKBERT P. S.: Surface simplification using quadric error metrics. In *Proc. SIGGRAPH* (1997), pp. 209–216. 3

[Hor18] HORN E. V.: *3D Character Development Workshop: Rigging Fundamentals for Artists and Animators*. Mercury Learning and Information, 2018. 1

[KLF11] KIM V. G., LIPMAN Y., FUNKHOUSER T.: Blended intrinsic maps. *ACM Trans. Graph. 30*, 4 (July 2011), 79:1–79:12. 4

[LMB14] LOPER M. M., MAHMOOD N., BLACK M. J.: MoSh: Motion and shape capture from sparse markers. *TOG 33*, 6 (Nov. 2014), 220:1–220:13. 1, 2

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graph. 34*, 6 (2015), 248:1–248:16. 2

[LRF10] LIPMAN Y., RUSTAMOV R. M., FUNKHOUSER T. A.: Biharmonic distance. *TOG 29*, 3 (2010), 27:1–27:11. 3

[LRR*17] LITANY O., REMEZ T., RODOLÀ E., BRONSTEIN A., BRONSTEIN M.: Deep functional maps: Structured prediction for dense shape correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017), IEEE, pp. 5659–5667. 4

[MMRC18] MARIN R., MELZI S., RODOLÀ E., CASTELLANI U.: Farm: Functional automatic registration method for 3d human bodies. *arXiv preprint arXiv:1807.10517* (2018). 2, 4, 7

[MOR*18] MELZI S., OVSJANIKOV M., ROFFO G., CRISTANI M., CASTELLANI U.: Discrete time evolution process descriptor for shape analysis and matching. *TOG 37*, 1 (Jan. 2018), 4:1–4:18. 3, 8

[MRCB16] MELZI S., RODOLÀ E., CASTELLANI U., BRONSTEIN M. M.: Shape analysis with anisotropic windowed fourier transform. In *2016 Fourth International Conference on 3D Vision (3DV)* (Oct 2016), IEEE, pp. 470–478. 3, 8

[MST*19] MELZI S., SPEZIALETTI R., TOMBARI F., BRONSTEIN M., DI STEFANO L., RODOLÀ E.: Gframes: Gradient-based local reference frame for 3d shape matching. In *IEEE Conference on Computer Vision and Pattern Recognition* (2019), CVPR, IEEE. 3, 8

[NO17] NOGNENG D., OVSJANIKOV M.: Informative descriptor preservation via commutativity for shape matching. *Computer Graphics Forum 36*, 2 (May 2017), 259–267. 4

[OBCS*12] OVSJANIKOV M., BEN-CHEN M., SOLOMON J., BUTSCHER A., GUIBAS L.: Functional maps: a flexible representation of maps between shapes. *TOG 31*, 4 (2012). 3, 4

[OCB*17] OVSJANIKOV M., CORMAN E., BRONSTEIN M., RODOLÀ E., BEN-CHEN M., GUIBAS L., CHAZAL F., BRONSTEIN A.: Computing and processing correspondences with functional maps. In *ACM SIGGRAPH 2017 Courses* (2017), pp. 5:1–5:62. 3, 4

[PLB12] PEARS N., LIU Y., BUNTING P.: *3D Imaging, Analysis and Applications*. Springer, 2012. 1

[PP93] PINKALL U., POLTHIER K.: Computing Discrete Minimal Surfaces and their Conjugates. *Exp. math. 2*, 1 (1993), 15–36. 2

[PSR*16] PICKUP D., SUN X., ROSIN P. L., MARTIN R. R., CHENG Z., ET AL.: Shape retrieval of non-rigid 3d human models. *International Journal of Computer Vision 120*, 2 (Nov 2016), 169–193. 2

[PWH*17] PISHCHULIN L., WUHRER S., HELTEN T., THEOBALT C., SCHIELE B.: Building statistical shape spaces for 3d human modeling. *Pattern Recognition 67* (2017), 276 – 286. 2

[RDP99] ROBINETTE K. M., DAANEN H., PAQUET E.: The caesar project: a 3-d surface anthropometry survey. In *Proc. 3DIM* (oct 1999), IEEE, pp. 380–386. 2

[RPWO18] REN J., POULENARD A., WONKA P., OVSJANIKOV M.: Continuous and orientation-preserving correspondences via functional maps. *ACM Trans. Graph. 37*, 6 (Dec. 2018), 248:1–248:16. 4

[Rus07] RUSTAMOV R. M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proc. SGP* (Aire-la-Ville, Switzerland, 2007), Eurographics Association, pp. 225–233. 3, 8

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L.: A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum 28*, 5 (2009), 1383–1392. 3, 8

[TSDS10] TOMBARI F., SALTI S., DI STEFANO L.: Unique signatures of histograms for local surface description. In *Proc. ECCV* (Berlin, Heidelberg, 2010), Springer-Verlag, pp. 356–369. 3, 8

[XZC18] XU Z., ZHANG Q., CHENG S.: Multilevel active registration for kinect human body scans: from low quality to high quality. *Multimedia Systems 24*, 3 (Jun 2018), 257–270. 2

[YBZW14] YAN D.-M., BAO G., ZHANG X., WONKA P.: Low-resolution remeshing using the localized restricted voronoi diagram. *TVCG 20*, 10 (2014), 1418–1427. 4

[ZSG*18] ZOLLHÖFER M., STOTKO P., GÖRLITZ A., THEOBALT C., NIESSNER M., KLEIN R., KOLB A.: State of the Art on 3D Reconstruction with RGB-D Cameras. *CGF 37*, 2 (2018). 1