

Geodesic Mapping for Dynamic Surface Alignment

Tony Tung, *Member, IEEE*, and Takashi Matsuyama, *Member, IEEE*

Abstract—This paper presents a novel approach that achieves dynamic surface alignment by geodesic mapping. The surfaces are 3D manifold meshes representing non-rigid objects in motion (e.g., humans) which can be obtained by multiview stereo reconstruction. The proposed framework consists of a geodesic mapping (i.e., geodesic diffeomorphism) between surfaces which carry a distance function (namely the global geodesic distance), and a geodesic-based coordinate system (namely the global geodesic coordinates) defined similarly to generalized barycentric coordinates. The coordinates are used to recursively choose correspondence points in non-ambiguous regions using a coarse-to-fine strategy to reliably locate all surface points and define a discrete mapping. Complete point-to-point surface alignment with smooth mapping is then derived by optimizing a piecewise objective function within a probabilistic framework. The proposed technique only relies on surface intrinsic geometrical properties, and does not require prior knowledge on surface appearance (e.g., color or texture), shape (e.g., topology) or parameterization (e.g., mesh connectivity or complexity). The method can be used for numerous applications, such as visual information (e.g., texture) transfer between surface models representing different objects, dense motion flow estimation of 3D dynamic surfaces, wide-timeframe matching, etc. Experiments show compelling results on challenging publicly available real-world datasets.

Index Terms—Geodesic mapping, surface alignment, dynamic surface, non-rigid deformation, 3D video, MRF.

1 INTRODUCTION

FREE-viewpoint video (i.e., 3D video) representing real humans or objects in motion immersed in a virtual world is becoming popular thanks to recent progresses in 3D reconstruction from multiple view videos (see [1]). As no markers or any special (wearable) equipment is necessary to perform the captures, the technique suits to a very wide range of applications such as surveillance, medicine, sports, entertainment, behavior understanding [2], etc.

Actually, 3D video consists of a stream of textured surface models undergoing free-form deformations and represented by 3D manifold meshes. As each 3D video frame is usually reconstructed independently using multiview stereo techniques [3], surface matching and tracking problems are not trivial while they are crucial for many of the applications cited above. To date, several solutions exist that: 1) use surface visual appearance (e.g., color or texture [4], [5]), or 2) perform matching by unilateral surface deformations [6].

We present a novel technique that achieves alignment of dynamic non-rigid surfaces using geodesic mapping (i.e., geodesic diffeomorphism). The geodesic mapping is defined on Riemannian manifolds which carry a distance function (the *global geodesic distance*) and a coordinate system (the *global*

geodesic coordinates). The proposed algorithm involves a coarse-to-fine strategy, where correspondences are found by searching for points with the same geodesic distances to landmark points; the set of landmark points being recursively updated with new correspondence points, chosen in specific regions, to increase the robustness to non-rigid deformations. When the distribution of correspondence points is dense enough, a mapping can be derived using generalized barycentric coordinates defined with respect to those points. Complete point-to-point surface alignment with smooth mapping is then obtained by optimizing a piecewise objective function within a probabilistic framework using a Markov Random Field (MRF) model. Our approach relies on surface intrinsic geometrical properties, and does not require prior knowledge on surface appearance (e.g., color or texture), shape (e.g., topology) or parameterization (e.g., mesh connectivity or complexity). Hence alignment accuracy is not lost in poorly textured regions (e.g., due to lack of texture, or poor photoconsistency) and can be achieved without surface remeshing (i.e., intrinsic geometrical properties are preserved). Several applications are shown in this paper: texture transfer, dense motion flow estimation, wide time frame matching, as well as surface alignment between models from different sequences. The rest of the paper is organized as follows. The next section discusses work related to the techniques involved in this paper. Section 3 presents our geodesic mapping strategy. Section 4 describes the 3D surface matching process. Section 5 shows experimental results. Section 6 concludes with a discussion on our contributions.

• The authors are with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. T.Tung is also with the Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606-8501, Japan.
Contact: www.tonytung.org, tm@i.kyoto-u.ac.jp

2 RELATED WORK

Several multi-view video capture systems have been developed since the 90s (see [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]). Usually, several sensing devices are spaced around a scene (e.g., in a studio) and synchronously perform the capture. The devices can either be a set of calibrated video cameras, or even handheld depth cameras. Additionally, 3D laser scanner can be used to leverage the reconstruction accuracy. With these techniques, subjects are captured without wearing any special markers, as opposed to motion capture methods (mocap). The resulting performance capture or 3D video consists of a stream of textured surface mesh models undergoing free-form deformation. Similarly, scientific interests related to 3D dynamic surface matching and tracking techniques have grown as many applications can indeed be derived, such as 3D motion flow estimation, temporal correspondence finding, labeling, surface deformation, spatio-temporally coherent reconstruction, etc.

In the past, the literature has provided several methods directly related to dense non-rigid surface matching and tracking. In general, the process consists of: (1) finding a sparse set of correspondence points between surfaces and (2) perform dense matching by regularization or mesh deformation. The correspondence problem is indeed particularly well-known for surface mesh edition and morphing (see [17] for a survey). Nevertheless it often requires that the user interactively defines numerous pairs of markers on the surfaces, which can be inaccurate and impracticable when the number of surfaces to match is high (e.g., like in 3D video). Hence, techniques for automatic surface feature extraction and matching are still actively under investigation.

Appearance-based approach. In [18], 3D scene flows are presented as an extension of optical flows [19] to represent the 3D motion of points. Optical flows are extracted and regularized in images and then used to derive the scene flows (that give point-to-point surface matching). In [20], scene flows and Laplacian mesh editing [21] are combined to track the deformation of high-quality a priori shape model obtained by laser-scan. In [4], the problem is addressed by combining Speeded Up Robust Features (SURF) [22] and Laplacian diffusion to obtain a dense displacement field. To guarantee the correctness of resulting meshes, surface morphing is applied as a final step. However the main limitation with these methods is the inability to capture the true deformation of low-frequency surface details (e.g., fingers, wrinkles in clothes, etc.) due to the implicit smoothing introduced by the mesh deformation process.

In [23], temporal correspondences are estimated by spherical parameterization and matching (but only for surfaces of genus-0). In [12], the authors

propose to use a set of surface descriptors (color feature, corner descriptor, edge descriptor and region descriptor) to derive a sparse set of points and construct a locally isometric mapping. Sparse-to-dense surface correspondence is then achieved using a MAP-MRF formulation. The accuracy is reported to be in the order of 5-10cm. In [14], meshes are deformed over time by tracking photo-consistent surface patches and optimization processes. In [24], SIFT [25] are used as 2D features to obtain sparse correspondences between adjacent frames. Thereafter, dense correspondences are generated by mapping harmonic functions [26] associated to the sparse points. This approach assumes isometric deformations, which is not always verified with real-world data (e.g., with large cloths). In [5], 2D local features are extended to 3D feature detectors and descriptors to improve sparse matching of meshes. The descriptors can capture photometric properties as well as local geometric properties (such as mean curvatures). In [15], 2D features (e.g., SIFT) are initially used with 3D information to estimate rough rigid transformations between surfaces (from colored depth maps). A mesh warping algorithm based on linear mesh deformation is then used to align the different partial surfaces. In [27], dense point clouds from scanner data are registered using a randomized feature matching algorithm relying on geodesics. As observed, the advantages of using local features such as SIFT, SURF and corners are the detection accuracy and robustness whereas local geometric features are more subject to surface noises due to reconstruction artifacts. Nevertheless, photometric feature matching approaches require surface models with good texture and color consistency between the multiple camera views, which is in practice tedious to obtain. Moreover, high resolution cameras are necessary to achieve dense matching, and using photometric features usually bounds the matching to surfaces having a unique texture (i.e., lighting cannot vary).

Geometry-based approach. Geometry-based approaches do not involve photometric features. In [6], a patch-based approach is used to track dynamic surfaces by unilateral deformations. However, as it implies a remeshing of the sequence with patches, intrinsic geometric properties (e.g., surface topology) and local geometric details can be lost. In [28] the authors propose the generalized multidimensional scaling (GMDS) that measures geodesics distortions in embedding spaces using Gromov-Hausdorff distances. GMDS is optimal for near-isometry deformations and average size surface meshes. Besides, a multi-scale process is necessary to avoid local minima during optimization computation. In [29], [30], [31], [32], the authors propose to use different geometry diffusion techniques, other than geodesics. However,

we show in Sect. 5 that these diffusion-based techniques are not as good as the proposed geodesic mapping to distinguish surface differences when the differences are very small.

In this paper, we present a novel approach for non-rigid (dynamic) surface alignment using geodesic mapping. A preliminary version of this work was presented in [33]. In particular, our solution is designed for 3D video data representing real-world performance (i.e., non synthetic) where surfaces can locally undergo non-rigid deformation and can contain reconstruction artifacts (i.e., surface noise, etc.).

Significant differences can be found compared to prior work on 3D surface registration that involves geodesics. In approaches like [34], [35], initial corresponding points based on local geometry properties (e.g., Euclidean distance or normal orientations) have to be determined beforehand. Hence, these methods cannot be well generalized to real-world 3D video data because of the surface noise inherent to the 3D reconstruction process. Furthermore in [34], geodesics are solely used to interpolate the positions of additional corresponding points by iteratively selecting mid-points of geodesic paths joining the points. This hierarchical scheme can only lead to surface alignment approximation, and no diffeomorphism (i.e., smooth map) is defined. Besides, an even distribution of initial points on the surfaces cannot be guaranteed on complex shape. In [35], a geodesic-based constraint is used for initial correspondence pruning. Additional correspondences are derived by propagation in the point neighborhoods using a global geodesic-based distortion measure similar to [28]. This method yields good result on synthetic and piecewise rigid surfaces, but again requires good initial corresponding points and is sensitive to surface noise, as there is no mechanism to ensure geodesic consistency between corresponding surface points. Moreover, the algorithm is applied only to reduced sample sets for efficiency reason and correspondence uniqueness is not handled. Thus, no mapping is explicitly defined.

In [36], [37], the authors perform non-rigid registration of depth scans using an ICP-based approach. The current formulation is specifically designed for partial matching of depth scans and yields good results for depth data which are initially closely related. The optimization consists in solving an objective function assuming a global rigid transformation using non-linear least squares method. (It can handle topology changes but has failure cases with wide-timeframe matching.) However, the convergence is not always guaranteed for full 3D surface model. The global minimization can be trapped by local minima and return locally large errors. Actually in [38] and later in [6], the authors applied an ICP-based method to full 3D models in order to track surface points of a deformable surface across time. Nevertheless, convergence failure can be observed (see Sect. 5.3). A

template is deformed across time (i.e., surface meshing is lost and topology is fixed), and the method is computationally expensive.

3 GEODESIC MAPPING

We propose to model a geodesic mapping between non-rigid surfaces by establishing a new framework which carries a distance (namely the global geodesic distance), and a coordinate system (namely the global geodesic coordinates). Dense sets of correspondence points are used to define geodesic-based coordinates (in the same fashion as generalized barycentric coordinates), while being recursively determined in non-ambiguous regions to robustly locate all surface points.

3.1 Problem formulation

Definition 1. Let us define two Riemannian manifolds (\mathcal{S}_1, g_1) and (\mathcal{S}_2, g_2) , where g_1 and g_2 are distance functions carried by \mathcal{S}_1 and \mathcal{S}_2 . The function $\Phi : \mathcal{S}_1 \rightarrow \mathcal{S}_2$ is a geodesic map (or geodesic diffeomorphism) if and only if:

- 1) Φ is a diffeomorphism of \mathcal{S}_1 onto \mathcal{S}_2 (i.e., a smooth bijective map),
- 2) Φ preserves geodesics (i.e., the image under Φ of a geodesic arc in \mathcal{S}_1 is a geodesic arc in \mathcal{S}_2),
- 3) Φ^{-1} preserves geodesics.

Assuming Φ defines a surface mapping between \mathcal{S}_1 and \mathcal{S}_2 , g_1 and g_2 return geodesic distances on \mathcal{S}_1 and \mathcal{S}_2 respectively.

To evaluate the surface mapping (or alignment) under a diffeomorphism $\phi : \mathcal{S}_1 \rightarrow \mathcal{S}_2$, one can define a metric $\delta : \mathcal{S}_1 \times \mathcal{S}_2 \rightarrow \mathbb{R}^+$, so that if $\exists \varepsilon \ll 1 : \forall v_1 \in \mathcal{S}_1, \delta(v_1, \phi(v_1)) \leq \varepsilon$, then $\phi(\mathcal{S}_1) = \mathcal{S}_2$. Ideally, our goal is to find Φ such as:

$$\Phi = \arg \min_{\phi} \left(\sum_{v_1 \in \mathcal{S}_1} \delta(v_1, \phi(v_1)) \right). \quad (1)$$

For example, surface alignment can be measured by evaluating geodesic distortions through an embedding (see [28], [30]). In this paper, we introduce the *global geodesic distance* D_N as a metric to define a geodesic mapping Φ that achieves non-rigid surface alignment without explicit embedding.

3.2 Global geodesic distance

3.2.1 Geodesic consistency

Definition 2. Assuming a set of N points $\mathcal{B} = \{b_1, \dots, b_N\}$ defined on a 2-manifold (\mathcal{S}, g) , the points v_1 and v_2 on \mathcal{S} are said to be geodesically consistent with respect to \mathcal{B} if and only if:

$$\forall i \in \{1, \dots, N\}, \quad |g(v_1, b_i) - g(v_2, b_i)| \leq \varepsilon', \quad (2)$$

where $\varepsilon' \ll 1$. If the points in \mathcal{B} do not have any particular configuration of alignment or symmetry,

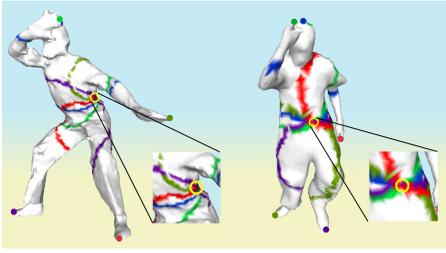


Fig. 1. Surface points (yellow circles) located at intersections of isolines (colored bands) given by geodesic distances to head and limb extremities.

the geodesic consistency property can be used to uniquely locate points on S when $N > 2$. In practice, uniqueness is verified by checking the number of iso-value line intersections from \mathcal{B} (see Fig. 1).

3.2.2 Global geodesic coordinates

Definition 3. Assuming a set of N points $\mathcal{B} = \{b_1, \dots, b_N\}$ defined on a 2-manifold (S, g) . For all $v \in S$, v can be located by its *global geodesic coordinates* with respect to \mathcal{B} , where the components are the N geodesic paths $\{g(v, b_1), \dots, g(v, b_N)\}$. The global geodesic coordinates can be seen as generalized barycentric coordinates defined on a Riemannian manifold. Hence, if S is a deformable surface and \mathcal{B} are fixed landmarks, the global geodesic coordinates can be used to locate surface points across time while S undergoes deformations. Besides, each component $g(v, b_i)$ is made invariant to isometric transformations by normalization with respect to $\max_{v \in S} g(v, b_i)$.

3.2.3 Distance definition

Definition 4. Let us assume two manifold surfaces (S^t, g^t) and (S^{t+1}, g^{t+1}) , where g^t and g^{t+1} are geodesic distance functions defined on S^t and S^{t+1} respectively, and $\mathcal{B}^t = \{b_1^t, \dots, b_N^t\}$ and $\mathcal{B}^{t+1} = \{b_1^{t+1}, \dots, b_N^{t+1}\}$ two sets of points defined on S^t and S^{t+1} respectively. We define the global geodesic distance $D_N : S^t \times S^{t+1} \rightarrow \mathbb{R}^+$ with respect to \mathcal{B}^t and \mathcal{B}^{t+1} as follows:

$$\begin{aligned} \forall (v, v') \in S^t \times S^{t+1}, \\ D_N(v, v') = \left(\sum_{i=1}^N |g^t(v, b_i^t) - g^{t+1}(v', b_i^{t+1})|^n \right)^{\frac{1}{n}}, \end{aligned} \quad (3)$$

where $n = 2$ (i.e., quadratic distance). Hence, assuming S^t undergoes temporal non-rigid deformations and \mathcal{B}^t are landmarks or surface points that are tracked over time such as critical points (see below), (v, v') forms a pair of matching points (namely correspondence points) if and only if $D_N(v, v') \leq \varepsilon''$, where $\varepsilon'' \ll 1$ is introduced to cope with distance approximations.

3.2.4 Initial feature extraction

Critical points. Let us assume that dynamic surfaces representing non-rigid objects in motion can be approximated by compact 2-manifold meshes. We choose geodesic distances to characterize surface intrinsic properties, as geodesic distances are invariant to pose (i.e., rotation and translation), and robust to shape variations (e.g., scaling) when normalized (see [41], [28]). Let $\mu : S \rightarrow \mathbb{R}$ denote the continuous function defined on the object surface S :

$$\mu(v) = \int_S g(v, s) dS, \quad (4)$$

where $g : S^2 \rightarrow \mathbb{R}$ is the geodesic distance between two points on S . Eq. 4 is the geodesic integral function whose critical points can be used to characterize shape (see Morse theory [42], [43]). The function μ is normalized with respect to its minimal and maximal values μ_{\min} and μ_{\max} as $\mu_N : S \rightarrow [0, 1]$, where $\mu_N(v) = \frac{\mu(v) - \mu_{\min}}{\mu_{\max} - \mu_{\min}}$. Maximal values of μ_N usually correspond to limb extremities (e.g., of human or animal models) while global minimum corresponds to body center. As illustrated in Fig. 2a, we can use μ_N to build Reeb graphs in order to identify and match critical points over time using geometry and topology information. Reeb graphs are high level shape descriptors that can be used for shape matching and retrieval in large datasets (see [41], [40], [44], [45]).

The graph construction relies on surface critical points, while the graph structure captures surface topology. Various local, global, geometry or topology information can then be embedded in the graph nodes with different weights, and graph matching can be achieved by defining a similarity function between graph nodes. In our experiments with real-world datasets (see Sect. 5), each node contains information about relative surface area, μ_N values, topology (i.e., valence), and shape distribution as described in [40], [2]. In practice, we do not observe any ambiguity when matching nodes of graphs extracted from consecutive frames of 3D video sequence. As natural (human) shapes and poses are usually asymmetric, nodes can be easily discriminated. However, in the case of data containing symmetry (e.g., if two legs are exactly similar), node trajectory or model priors (e.g., on shape, orientation, color, etc.) can also be embedded to prevent ambiguity.

Note that global geodesic distances between matching critical points are usually small ($\leq \varepsilon''$) when the matchings are correct. For complex scenario of critical point tracking, one can refer to [46] and [47]. Thus, just as 3 points are sufficient to define a barycentric coordinate system, all vertices of S can be *uniquely* identified by global geodesic coordinates with $N \geq 3$ components. Uniqueness is discussed in Sect. 3.3.1.

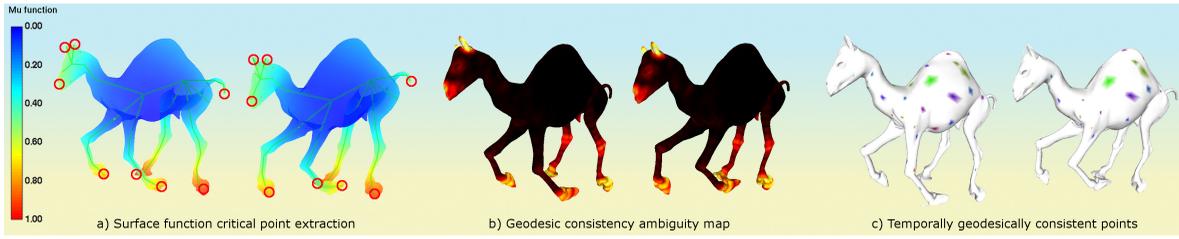


Fig. 2. a) Critical points extracted on two surface meshes from [39] using Reeb graphs [40] (red circles). b) Point localization ambiguity with respect to critical points (darker means less ambiguity). c) 50 correspondence points found between the two surfaces using global geodesic coordinates.

Robustness. 3D video data usually contain 3D models with reconstruction artifacts. Fortunately, the Morse function μ is robust to local surface noise thanks to the integral formulation on S which smooths out outlier geodesics. For additional discussions on robustness of μ , refer to [2]. In what follows, we present a coarse-to-fine strategy to determine robust correspondence points between dynamic surfaces by evaluating point localization ambiguity.

3.3 Correspondence point matching

3.3.1 Correspondence points

As presented in the previous sections, global geodesic coordinates can be used to match points located on deformable manifold surfaces. Assuming the set of M points $\{v_1^t, \dots, v_M^t\} \in S^t$, the correspondence problem consists of finding the geodesic map Φ such that:

$$\forall v^t \in S^t : \Phi(v^t) = \arg \min_{v^{t+1} \in S^{t+1}} (D_N(v^t, v^{t+1})), \quad (5)$$

where $D_N : S^t \times S^{t+1} \rightarrow \mathbb{R}^+$ defines a global geodesic distance with respect to a set of N local features (e.g., critical points) as introduced in Def. 4. The set of pairs $\{(v_1^t, v_1^{t+1}), \dots, (v_M^t, v_M^{t+1})\}$ where $v_i^{t+1} = \Phi(v_i^t), \forall i \in \{1, \dots, M\}$, forms the correspondence points between S^t and S^{t+1} (see Fig.2b).

However, when surfaces undergo non-rigid deformations (e.g., non-isometric), geodesic distances may not be preserved and global geodesic coordinates may not be sufficient to uniquely determine a point position. Hence, we propose to adopt a coarse-to-fine strategy to recursively choose correspondence points in regions where position ambiguity is low.

3.3.2 Ambiguity estimation

As mentioned in the previous sections, non-rigid deformations can affect geodesics consistency. As a consequence, surface point localization can become ambiguous.

Definition 6. Assuming a manifold S and a set of N points $\mathcal{B} \in S$, we can estimate point localization ambiguity on S with respect to \mathcal{B} by defining a local ambiguity degree $\mathcal{A}(v)$, $\forall v \in S$, which is obtained as

the measure of the number of points geodesically consistent to v with respect to \mathcal{B} (see Sect. 3.2.1). When the ambiguity degree is computed for all points $v \in S$, we can represent $\{\mathcal{A}(v)\}$ as an *ambiguity map* (see Fig. 2c).

However, when \mathcal{B} contains points located only at distant surface extrema (e.g., extremities of human limbs), one can observe that regions located around the object center have very low ambiguity degree (i.e., geodesic consistency uniqueness can be established), while regions around the extrema are more ambiguous. Nevertheless, ambiguity issues can be solved by: 1) increasing N , and 2) having points of \mathcal{B} closer to regions of high ambiguity degree.

3.3.3 Recursive mapping

To guarantee that all surface point positions can be determined without ambiguity, we propose to recursively populate \mathcal{B} with points located in regions with low ambiguity. When applying this strategy simultaneously for two surfaces S^t and S^{t+1} , and recursively populating the sets $\mathcal{B}^t \in S^t$ and $\mathcal{B}^{t+1} \in S^{t+1}$ with pairs of correspondence points (see Sect. 3.3.1), we obtain a discrete geodesic map Φ_N (i.e., using N landmarks) of S^t onto S^{t+1} . Figure 3 illustrates the recursive (coarse-to-fine) mapping process. The algorithm is as follows:

- 1) Initialize \mathcal{B}^t and \mathcal{B}^{t+1} with sets of local features that can be matched and tracked (e.g., critical points as described in Sect. 3.2.4).
- 2) Compute ambiguity maps for S^t and S^{t+1} with respect to \mathcal{B}^t and \mathcal{B}^{t+1} respectively.
- 3) Choose randomly N_i pairs of correspondence points between S^t and S^{t+1} , where both points in each pair are located in regions with low ambiguity with respect to \mathcal{B}^t and \mathcal{B}^{t+1} respectively. Points are chosen to be spaced by at least $\sqrt{\frac{\text{area}(S^t)}{(N+N_i)\pi}}$. In practice, correspondence points are chosen among vertices with low ambiguity degree: $\{v \in S : \mathcal{A}(v) < \tau\}$, where $\tau = \min(\mathcal{A}(v)) + 0.1 * (\max(\mathcal{A}(v)) - \min(\mathcal{A}(v)))$.
- 4) Insert the $2N_i$ pair points in \mathcal{B}^t and \mathcal{B}^{t+1} respectively.
- 5) Set $N = \sum_i N_i$.

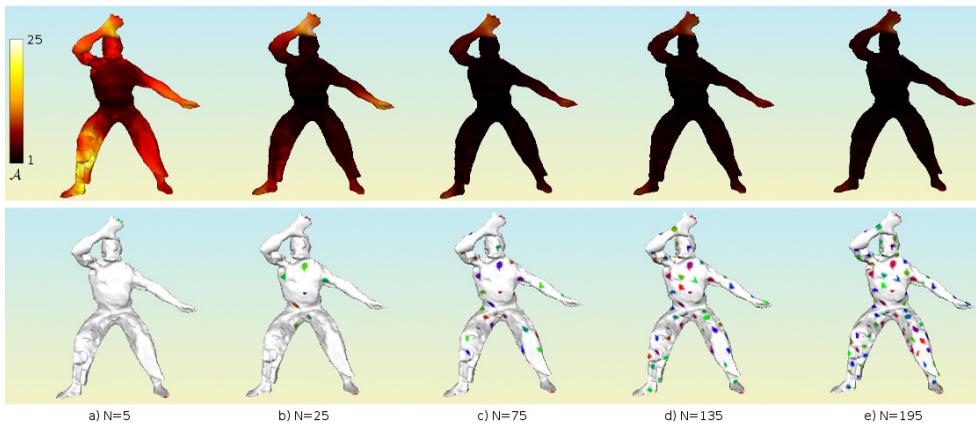


Fig. 3. Coarse-to-fine process to determine N correspondence points. Top) Ambiguity degree \mathcal{A} of each surface point with respect to N . Bottom) Correspondence points in non-ambiguous regions.

- 6) Repeat 2) to 5) until no regions with high ambiguity degree remain (e.g., $\forall v \in \mathcal{S} : \mathcal{A}(v) < 2$).

Finally, we obtain a mapping function $\Phi_N : \mathcal{S}^t \rightarrow \mathcal{S}^{t+1}$ that relates surface points which are geodesically consistent and should ideally represent a geodesic mapping. However, if surfaces are modeled by manifold meshes reconstructed independently (as with 3D video data), surface mesh connectivity and number of vertices can be different. A real bijection between the surface models is then impracticable to realize, as mesh edge lengths would have to be smaller than ε . Hence, we consider Φ_N as a surjective function where each point position in \mathcal{S}^{t+1} is approximated by the nearest vertex in the corresponding mesh model. This choice is reasonable when the vertex number is high, and guarantees that all surface point of \mathcal{S}^t is matched.

Nevertheless, with large meshes (e.g., 300,000 vertices), the process can become computationally expensive as N can become very high (the iteration steps 2) and 3) have quadratic complexity). Thus, we use only around $N = 200$ correspondence points. Complete surface alignment is achieved by propagation as described in the next section. For example, Fig. 4 presents the ratio of non-ambiguous surface regions with respect to the number N of correspondence points computed on the sequence Crane (obtained from real-word capture [48]). When the ratio is 100%, then $\forall v \in \mathcal{S} : \mathcal{A}(v) < \tau$.

4 NON-RIGID SURFACE ALIGNMENT

We present an MRF energy formulation to achieve dense and point-to-point smooth alignment of non-rigid manifold surfaces, where the correspondence points defined in the previous section serve as priors.

4.1 MRF formulation

Let $\mathcal{P} = \{p_i\}$ be the set of vertex sites on a surface mesh \mathcal{S}_1 , and $\mathcal{L} = \{l_p\}$ be a discrete set of labels

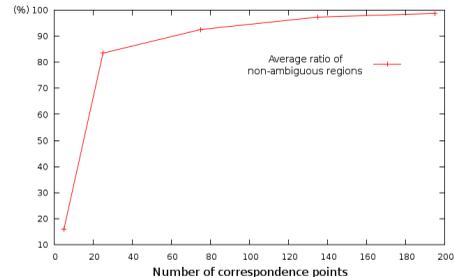


Fig. 4. Percentage of non-ambiguous surface points with respect to the number N of correspondence points.

corresponding to the candidate vertices on a surface mesh \mathcal{S}_2 . We propose to minimize the following MRF energy in order to achieve complete mapping of \mathcal{S}_1 onto \mathcal{S}_2 :

$$E(f) = E_d(f) + E_s(f). \quad (6)$$

$E(f)$ is the total energy of the labeling (or mapping) $f : \mathcal{P} \rightarrow \mathcal{L}$, $E_d(f)$ is the data term and $E_s(f)$ the smoothness term:

$$\begin{aligned} E_d(f) &= \lambda_d \cdot \sum_{p \in \mathcal{P}} D_N(p_i, l_p), \\ E_s(f) &= \lambda_s \cdot \sum_{\{p, q\} \in \mathcal{N}} |T_p(l_p) - T_q(l_q)|, \end{aligned} \quad (7)$$

where \mathcal{N} is a neighborhood configuration in \mathcal{S}_1 , $\lambda_d = 0.75$ and $\lambda_s = 0.25$ are a constant factors, and $T_p(l_p)$ and $T_q(l_q)$ are vector flows joining p to l_p , and q to l_q respectively. Actually, this is an NP-complete problem for which minimization is very computationally expensive using the usual techniques [49], [50].

4.2 Divide and conquer

We propose to dramatically reduce the label search space by dividing the global optimization problem into smaller local ones, and consequently make the optimization process tractable. This is valid as the priors (i.e., the correspondence points) are independent

of each other. The strategy consists of sweeping the surface mesh using front propagations from an initial point in \mathcal{B} located near the center of the object (e.g., given by $\min_{v \in \mathcal{S}}(\mu_N(v))$), propagate the mesh vertex matching to neighbor regions (i.e., the closest correspondence points), and process unmatched vertices.

Let us denote $\mathcal{N}(p)$ the neighborhood of $p \in \mathcal{S}_1$ and $\mathcal{N}(l_q)$ the neighborhood of $l_q \in \mathcal{S}_2$, and let us assume the local continuity property of diffeomorphism: $\forall q \in \mathcal{N}(p), \exists l_p \in \mathcal{N}(l_q)$. Hence, for every vertex site p : 1) we can look for the closest (already matched) correspondence points $\{q\} \subset \mathcal{N}(p)$, and 2) find l_p in $\bigcup_{\{q\} \subset \mathcal{N}(p)} \mathcal{N}(l_q)$ by minimizing the energy as defined in Eq. 6. Note that in our experiments, we use two-ring neighborhoods (i.e., 15 to 20 vertices).

Figure 5 illustrates surface mapping between two surface models. In a), we show piecewise mapping using $N = 195$ correspondence points (and their neighborhood). In b), we show complete surface mapping after optimization. The color codes of surfaces represent global geodesic coordinates using 3 reference points (located on the head and at the feet of the models) encoded as RGB color channels. Note that here, local geometry-based descriptors (e.g., [5]) cannot be used to find correspondence points as the models contain too much surface noise.

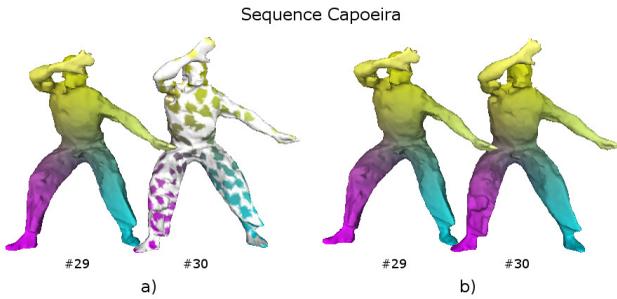


Fig. 5. Surface mapping between two models (left and right) before and after optimization: a) 195 correspondence points found on the model on the right give piecewise mapping. b) After optimization, we obtain complete surface alignment.

4.3 Implementation details

4.3.1 Non-rigid deformation management

Although global geodesic coordinates are defined with respect to landmarks which are chosen in non-ambiguous regions, some coordinate components can be unreliable if their corresponding geodesic paths pass through surface regions undergoing large non-rigid deformations. As a consequence, the set of candidates among which to choose the correspondence points as presented in Sect. 3.3.2 can remain large (i.e., $\mathcal{A}(v) > 1$), and the landmark positions can be unoptimally defined (e.g., the most central candidate is not always the best).

Hence, we introduce additional conditions at step 3) in Sect. 3.3.3 to avoid false candidates. Assuming a set of N pairs of corresponding points $\{(b_1^1, b_1^2), \dots, (b_N^1, b_N^2)\}$ in $(\mathcal{S}_1, g_1) \times (\mathcal{S}_2, g_2)$, and a point $p \in \mathcal{S}_1$ to be matched, we find the matching point $q \in \mathcal{S}_2$ determined by:

- a maximum number N_{pq} of reliable pairs of coordinate components, i.e., where most isovalue lines intersect, as discussed Sect. 3.2.1,
- the minimum global geodesic distance to $p \in \mathcal{S}_1$: $q = \arg \min_{q' \in \mathcal{S}_2} D_{N_{pq'}}(p, q')$.

Hence, each candidate point q' has to comply with the following two conditions:

$$\forall i \in \{j_1, \dots, j_{N_{pq'}}\}, |g_1(p, b_i^1) - g_2(q', b_i^2)| < \varepsilon', \quad (8)$$

$$\text{and } D_{N_{pq'}}(p, q') < \varepsilon,$$

where $\{j_1, \dots, j_{N_{pq'}}\} \in \{1, \dots, N\}$ are $N_{pq'}$ indices of reliable coordinate components verifying Eq. 8.

Furthermore, the positions of the correspondence points $\{b_1^2, \dots, b_N^2\}$ in \mathcal{S}_2 are globally optimized after each iteration of the recursive mapping described in Sect. 3.3.3 to correct eventual error propagations as follows:

$$\{b_1^2, \dots, b_N^2\} = \arg \min_{b_i^2 \in \mathcal{N}(b_i^1)} \sum_{i=1}^N D_{N_{b_i^1 b_i^2}}(b_i^1, b_i^2), \quad (9)$$

where $\mathcal{N}(b_i^1)$ is the neighborhood of b_i^1 (e.g., two-ring).

Hence, by choosing the number N of landmark high enough, inaccurate coordinate components become less significant. The assumption is that there always exists a subset of coordinate components that allows any pair of points $(p, q) \in \mathcal{S}_1 \times \mathcal{S}_2$ to remain geodesically consistent (i.e., $N_{pq'} > 2$). However, the continuity condition introduced in the previous section does not leave unmatched surface points.

For example, to illustrate the implementation details we reported in Table. 1 the average errors (i.e., distance to ground truth) when calculating the N correspondence points between successive frames of a 3D video sequence. We used the sequence Crane (173 frames) processed by [6] as ground truth, as each of the 3407 vertices is tracked across the sequence. Errors Err_1 and Err_2 correspond to the number of correspondence points $\{b_1^2, \dots, b_N^2\}$ which are not exactly correctly placed (i.e., b_i^1 and b_i^2 have different vertex indices), after each iteration of the recursive mapping, before and after the optimization process respectively (Eq. 9). Note that this is a very strict condition for error evaluation. As well, the errors of complete surface alignment Err_{tot} is given after each iteration. They correspond to the average distance to ground truth, and are expressed as a percentage of the mesh size.

As can be observed in Table 1 and Fig. 6, surface alignment accuracy is improved by the proposed recursive geodesic mapping, and reach near optimal mapping after few iterations ($Err_{tot} = 0.55\%$). Additional experimental results are presented in Sect. 5.

<i>iteration</i>	0	1	2	3	4
N	5	25	75	135	195
Err_1		8	30	47	57
Err_2		7	25	32	43
$Err_{tot}(\%)$	2.20	1.61	0.84	0.60	0.55

TABLE 1

Number of mismatches of N correspondence points before (Err_1) and after (Err_2) optimization (see Eq. 9), and average distance errors (Err_{tot}) after each iteration of the recursive mapping obtained on the sequence Crane (ground truth from [6]).

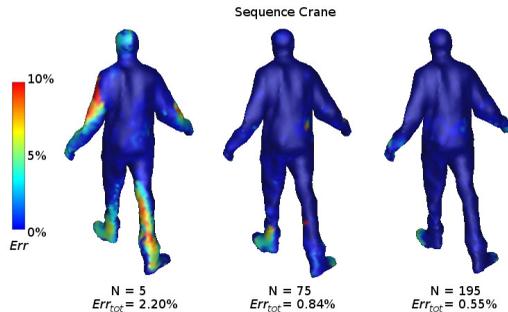


Fig. 6. Surface alignment errors with respect to the number of landmark points N . Err represents the distance to ground truth as a percentage of the mesh size. Err_{tot} is the average error over the whole 3D video sequence (sequence Crane contains 173 frames).

4.3.2 Topology change management

Topology-based approaches that involve geodesics computed over object surface are usually sensitive to surface topology changes. For example, when body parts touch each other (e.g., arms akimbo), the global distribution of geodesics is usually altered. Although it can be desirable for some applications (e.g., shape matching [51], scene understanding [2]), surface topology invariance is rather desirable in our context.

However, as presented in the previous sections, our algorithm relies on a piecewise mapping where the priors (i.e., correspondence points) are located in non-ambiguous regions. Particularly, the recursive process presented in Sect. 3.3.3 prevents surface point matching in ambiguous regions. As illustrated in Fig. 7, after few iterations, surface regions where no topology change occurs do not suffer from misalignment, and correspondence points can be found close to disturbed regions where topology changes occur. On the other hand, even a method using several conformal maps [31] cannot cope directly with topology change (see wrong surface matchings in green in Fig. 7d).

Note that in [6], the patch-based alignment method implicitly handles topology changes across 3D video sequence. However, the optimization process can locally converge to a wrong solution during few frames (see next Sect. 5). Besides, surface mesh connectivity

is lost (as surfaces are remeshed) as well as local geometry information (i.e., surface details are lost).

5 EXPERIMENTAL RESULTS

3D video datasets and ground truth. For surface alignment evaluation, we use several publicly available datasets of 3D video sequences, such as the sequences Kickup, Lock, Pop and Free from the University of Surrey [12], the sequences Bouncing, Crane, Handstand and Samba from MIT-CSAIL [48], and the sequences Yoga and Capoeira from Kyoto University [8]. In the sequences from [12], the subject wears a loose T-shirt and performs quick dance steps. Every mesh is high-resolution (140,000 vertices) and has a smooth surface which contains local reconstruction artifacts. The sequences from [48] have similar features, while the surfaces are smoother (i.e., less surface noise) but more rigid. The sequences from [8] contain high-resolution meshes (around 15,000 vertices) with more reconstruction artifacts caused by non-homogeneous lighting from different viewpoints and poorly textured surface regions.

These datasets are relevant as they are obtained from real human performances, and challenging as they contain complex surface motions of human body and cloths, as opposed to synthesized datasets which usually consist only of piecewise rigid surfaces (and usually serve for shape matching and retrieval).

In [6] the authors have proposed a patch-based surface tracking technique that is robust across time, and consistently remeshed the datasets cited above (all mesh models in each sequence share the same connectivity). Hence, we will consider the datasets provided by [6] as ground truth data, and discard from evaluations erroneous frames (identified by hand).

Computation cost. The algorithms were developed in C++ on standard machines (Corei7 CPU). Geodesic distances on surface mesh models are approximated by shortest paths on edges, and are computed using the Dijkstra's algorithm with a binary heap which complexity is $|V|\log|V|$, where $|V|$ is the number of vertices in a surface mesh. Although the current implementation has not been optimized for high computation time performance, the alignment process duration is reasonable even without GPU implementation (e.g., in the order of seconds with $|V| = 15,000$ if $N < 50$, and 1min if $N = 195$). Computation time between two surfaces is proportional to the size of both source and destination meshes (e.g., around 20s if $|V| = 30,000$ and $N = 75$). The most time consuming step is actually the computation of the geodesic integral (Eq. 4) which takes around 65% of the total calculation (see Fig. 8).

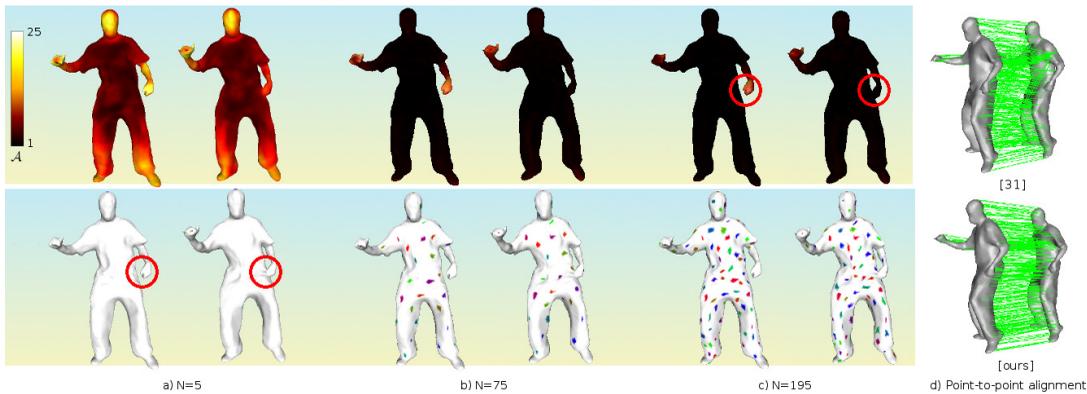


Fig. 7. Matching of N pairs of correspondence points between surfaces of different topology. The surface mesh on the right (in a), b) and c)) has a self-intersection as the left hand is touching the hip (see red circle). However, the iterative computation of non-ambiguous regions allows accurate matching of correspondence points (where $\mathcal{A} = 1$), except for the local region around the disturbed area. In d), our method is compared to [31] for point-to-point surface alignment (10% of matchings are represented).

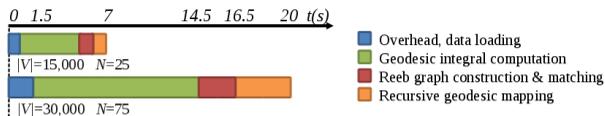


Fig. 8. Computational cost details.

Baseline. We evaluate point-to-point surface alignment against state-of-the-art surface alignment techniques [31], [32]. In [31], correspondences are obtained from multiple conformal maps that are blended together. This method (Blended Intrinsic Maps) is known to outperform several other approaches (e.g., [28]) on various datasets. In [32], the authors propose to perform matching in the spectral domain using spectral embeddings, as they are less dependent to rigid and non-rigid deformations. This method (FOCUSR) has been tested on various datasets and showed good performances. In our evaluations, we use FOCUSR with simple spectral coordinates (as described in [32]). As our approach, these methods do not need prior information, do not rely on color or texture, and do not depend on local geometry information. The latter point is essential as 3D video data can contain strong surface noise. As surface alignment can be used for numerous applications, we show some examples of wide-timeframe matching [52], texture transfer, and 3D motion flow estimation [4].

5.1 Surface alignment evaluation

Frame-to-frame alignment. We start our evaluation by showing that the proposed geodesic mapping is accurate for point-to-point surface alignment across 3D video sequences. Although the global shapes can look similar from frame to frame, the problem is actually not trivial when surfaces undergo non-rigid deformations. Figure 9 (top) illustrates point-

to-point surface alignment between frames of the sequence Bouncing. The sequence consists of 174 frames and contains large non-rigid surface deformations (a subject wears loose clothing and jumps around). Our method yields an average distance error of 1.18% ($\pm 0.75\%$) to ground truth (percentage calculated with respect to the mesh size), and the Blended Intrinsic Maps [31] return 1.05% ($\pm 0.58\%$). Both significantly outperforms the spectral method (FOCUSR with simple spectral coordinates) [32] which achieves 10.70% ($\pm 1.96\%$). Furthermore, as shown on the right, the patch-based tracking method [6] can converge to local optima and produce erroneous surface geometry. In that case, a complete mapping cannot be performed (see circle). In comparison, our method can be applied directly to original data. Similarly, Fig. 9 (bottom) illustrates point-to-point surface alignments between frames of the sequence Free (499 frames). Our method returns an average distance error of 1.35% ($\pm 1.03\%$), [31] returns 0.97% ($\pm 0.62\%$), and [32] achieves 10.05% ($\pm 1.45\%$). As well, the figure on the right shows erroneous convergence of [6], and we show the surface matching obtained on the original dataset using our method.

Supplemental evaluations on sequences Pop, Lock, Crane, Samba, Handstand and Kickup are presented in Fig. 10, where we show examples of alignments using [32] and our method. Note that we do not illustrate the results obtained with [31] as they are visually similar to ours. All average distance errors to ground truth are reported in Table. 2. Graphs representing comparisons of point-to-point alignment methods ([32], [31] and ours) computed between consecutive frames of sequences are shown in Fig. 11. In general, we observe that the spectral method [32] is not as accurate as the proposed geodesic mapping for point-to-point surface alignment across 3D video data, while [31] shows similar performances than our

Dataset	Free	Pop	Lock	Bouncing	Crane	Samba	Handstand	Kickup
#frames	499	249	249	174	173	174	173	219
[32] (%)	10.05(\pm 1.45)	6.69(\pm 0.64)	7.40(\pm 1.11)	10.70(\pm 1.96)	8.33(\pm 0.79)	4.93(\pm 0.66)	9.15(\pm 2.93)	8.68(\pm 0.94)
[31] (%)	0.97(\pm 0.62)	0.22(\pm 0.22)	0.64(\pm 0.55)	1.05(\pm 0.58)	0.58(\pm 0.24)	0.70(\pm 0.31)	0.63(\pm 0.46)	0.41(\pm 0.39)
[ours] (%)	1.35(\pm 1.03)	0.19(\pm 0.27)	0.73(\pm 1.74)	1.18(\pm 0.75)	0.55(\pm 0.28)	0.70(\pm 0.30)	0.75(\pm 0.55)	0.40(\pm 0.51)

TABLE 2

Average of relative distance errors from ground truth (with standard deviation) for surface alignments between consecutive frames of various 3D video sequences.

propose method. However, [31] relies on multiple conformal maps (whereas we use only one), and does not cope well with topology change (see Sect. 4.3.2 and Fig. 7d).

Wide-timeframe alignment. Figure 12 illustrates surface alignment for wide-timeframe matching (as in [52]). The process is also achieved without using any prior articulated shape model or frame-to-frame surface tracking, while they are the usual straightforward (non-optimal) solutions. Our approach returns complete alignment of surfaces, where all points are mapped. We ran several tests on the datasets and found our method performs much better than the spectral matching method [32] and the blended intrinsic maps [31] when they fail at correctly matching surface protrusions. For example alignment errors (\bar{Err}) between frames #29 and #125 of sequence Free are 38.16% for [32], 32.48% for [31], and 4.14% for our approach.

5.2 Applications

Texture transfer. As illustrated in Fig. 13, the accuracy of our approach allows us to perform texture transfer between dynamic surfaces. In a), texture transfer is performed between a model from the sequence Capoeira and a model from the sequence Free. The texture transfer returns globally good visual results although some minor local artifacts can be observed. Note that the wrinkles on the shirts are particularly well transferred. In b), a model from the sequence Yoga with real texture is matched with an untextured model, and texture transfer is performed efficiently (see front and back). In c), d), e) and f), various effective texture transfers are shown between textured mesh sequences from [6] (from the models on the left to the models on the right).

3D flow estimation. Motion flows are obtained by alignment of consecutive frames of 3D video data. Figure 14a) shows motion flows computed on the sequence Capoeira, where we can observe dense motion flows. Figure 14b) shows motion flows computed on two clips of the sequence Samba. Here, all surface points are mapped using geodesic mapping. As a comparison, in [4], the authors could not perform well on that sequence due to the poor

surface texture. Note that as no vertex tracking over time is performed directly, we do not have connectivity consistency between consecutive frames. Nevertheless, this could be naively obtained by choosing nearest neighbor flows.

5.3 Failure case

According to our experiments, despite overall good performance, mapping can be locally unperfect when severe variations of topology or shape occur. In Fig. 15, we show examples of alignments with surfaces containing errors found in some sequences of [6]. Unfortunately here, ground-truth is only partially correct. When reconstruction errors are not too severe, we can observe that surface alignment is globally correct (e.g., see Bouncing).

Note that, as our current implementation requires surface mesh edges to compute geodesics, a meshing or remeshing step is mandatory when applying the method directly to point clouds and partial scans. As well, we expect the surfaces to be aligned to have globally consistent shape. For example, matching human to gorilla is problematic because of the different proportions of limbs.

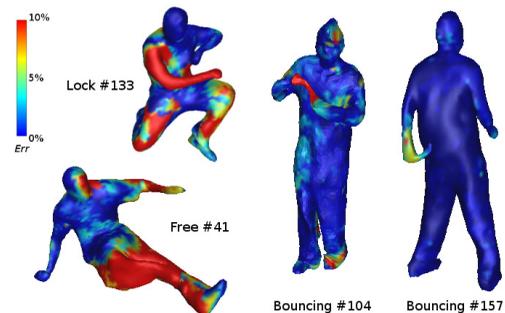


Fig. 15. Misalignments caused by strong variations in shape or topology.

6 CONCLUSION

In this paper, we present a novel approach to perform dynamic surface alignments. We define a geodesic mapping (i.e., diffeomorphism) between surfaces considered as Riemannian manifolds that carry a specific

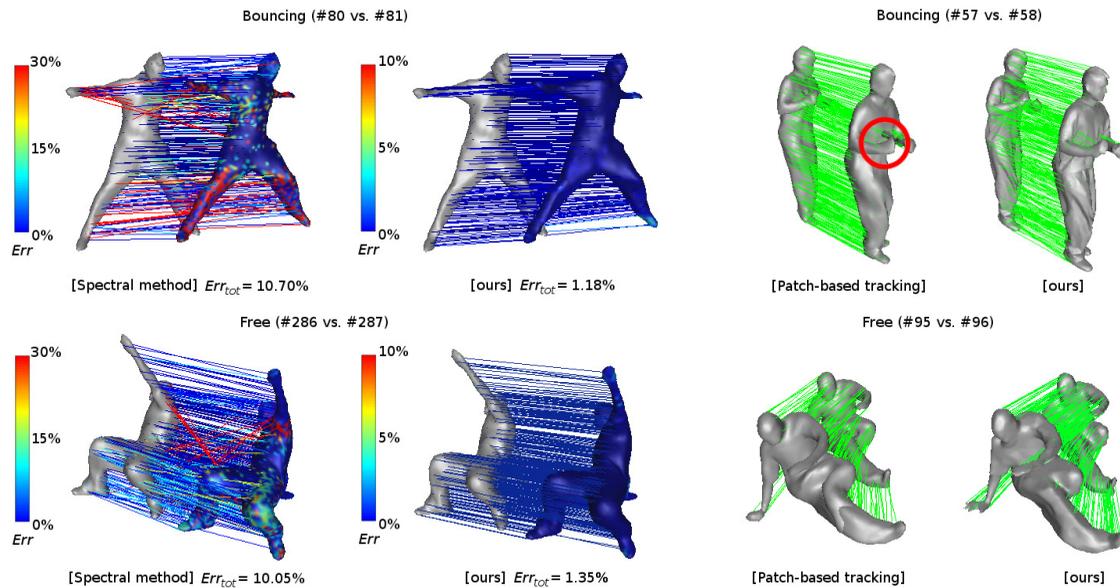


Fig. 9. Left) Surface alignment errors on the datasets Bouncing (top) and Free (bottom). Our approach outperforms the spectral method. Right) Patch-based tracking can return local reconstruction errors, whereas our matching method performs alignment without altering surface geometry. (10% of matchings are represented.)

distance function we call global geodesic distance. Particularly, as non-rigid deformations are known to affect geodesics, the morphism is built using a coarse-to-fine strategy where correspondence points between surfaces are sought in non-ambiguous regions. The correspondence points are recursively used to define the global geodesic coordinates (similar to generalized barycentric coordinates) that serve to uniquely locate surface points. Complete point-to-point surface mapping is then obtained by formulating the problem as a piecewise MRF optimization problem in order to make the problem tractable, and using the correspondence points as priors. Our approach does not require color-based feature extraction (unlike [4], [5]), nor surface remeshing or deformation (unlike [6]), and can handle very large meshes (unlike [28]). As well, sensitivity to topology change is discussed and we show that our approach is competitive compared to the state-of-the-art. Finally, this algorithm is used for numerous applications such as texture transfer, 3D motion flow estimation in 3D video, wide-timeframe matching, etc. The performances are obtained on high-resolution mesh models from publicly available real-world datasets.

For further improvements, we believe that color information could be used when available, and as well, different geometry diffusion methods could be explored to improve the robustness to topology variations.

ACKNOWLEDGMENTS

This work was supported in part by the JST-CREST project "Creation of Human-Harmonized Information

Technology for Convivial Society".

REFERENCES

- [1] T. Matsuyama, S. Nobuhara, T. Takai, and T. Tung, "3D video and its applications," *Springer*, 2012.
- [2] T. Tung and T. Matsuyama, "Topology dictionary for 3D video understanding," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 34, no. 8, pp. 1945–1957, 2012.
- [3] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.
- [4] K. Varanasi, A. Zaharescu, E. Boyer, and R. Horaud, "Temporal surface tracking using mesh evolution," *Proc. European Conf. Computer Vision*, 2008.
- [5] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.
- [6] C. Cagniard, E. Boyer, and S. Ilic, "Probabilistic deformable surface tracking from multiple videos," *Proc. European Conf. Computer Vision*, 2010.
- [7] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka, "A stereo machine for video-rate dense depth mapping and its new applications," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1996.
- [8] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara, "Real-time 3D shape reconstruction, dynamic 3D mesh deformation, and high fidelity visualization for 3D video," *Computer Vision Image Understanding (CVIU)*, vol. 96, no. 3, pp. 393–434, 2004.
- [9] J. Franco, C. Menier, E. Boyer, and B. Raffin, "A distributed approach for real-time 3D modeling," *CVPR Workshop on Real-Time 3D Sensors and their Applications*, p. 31, 2004.
- [10] K. M. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking," *Int'l Jour. Computer Vision (IJCV)*, vol. 63, no. 3, pp. 225–245, 2005.
- [11] J. Allard, C. Ménier, B. Raffin, E. Boyer, and F. Faure, "Grimage: Markerless 3D interactions," *SIGGRAPH - Emerging Technologies*, 2007.
- [12] J. Starck and A. Hilton, "Surface capture for performance-based animation," *IEEE Computer Graphics Applications*, 2007.

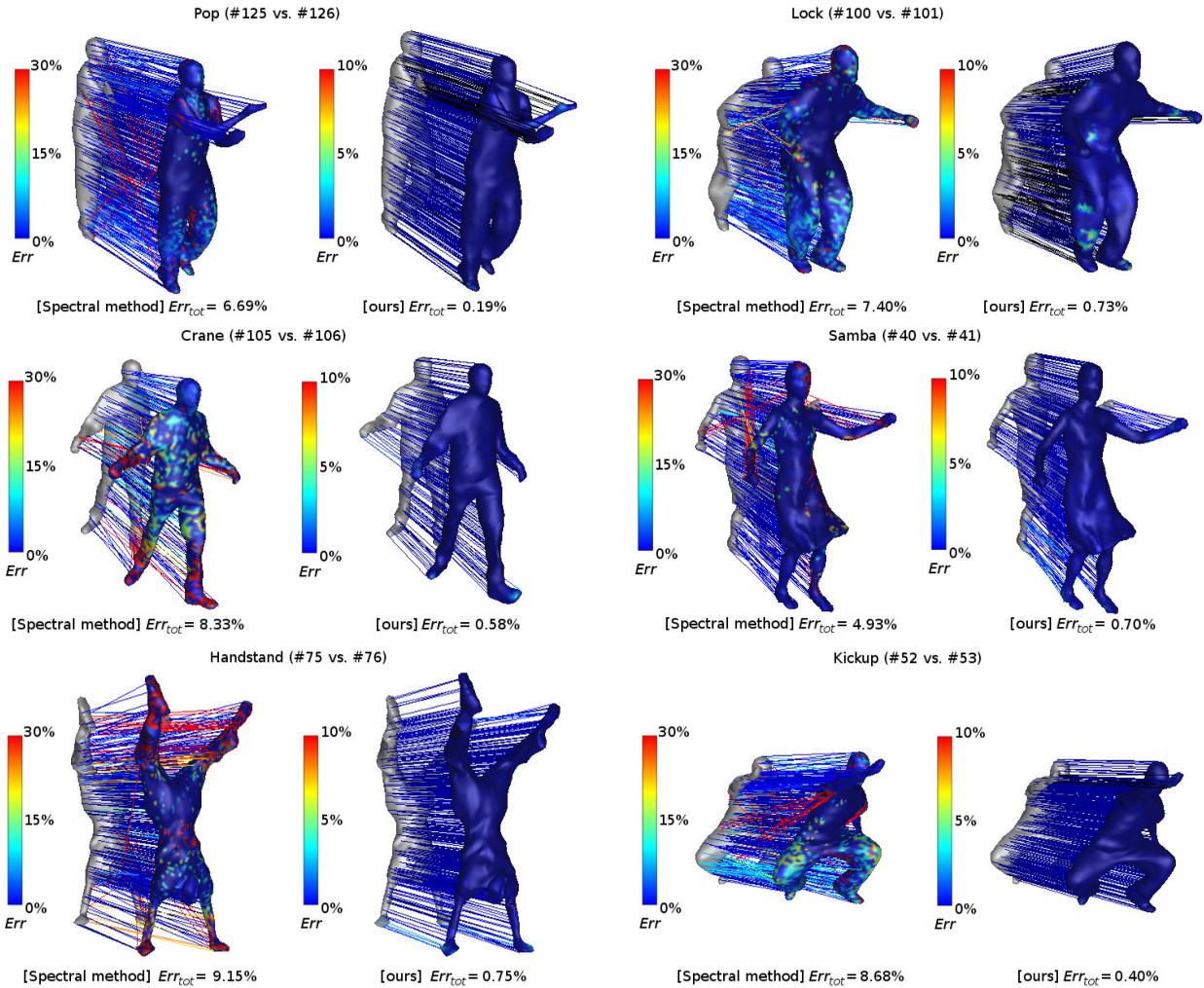


Fig. 10. Surface alignment errors on the datasets Pop, Lock, Crane, Samba, Handstand and Kickup. Our approach outperforms the spectral method. (10% of matchings are represented.)

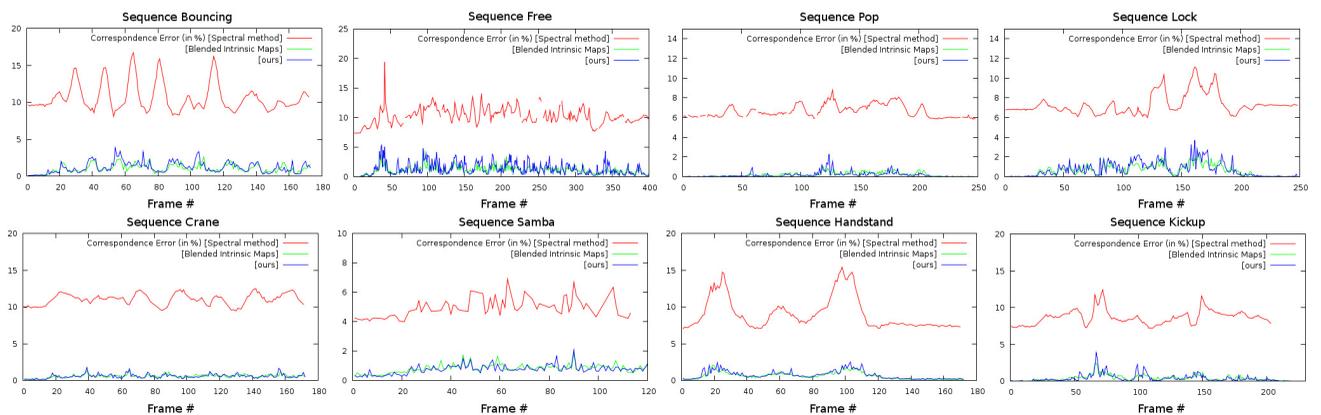


Fig. 11. Comparison of point-to-point alignment methods ([32], [31] and ours) computed between consecutive frames of sequences Bouncing, Free, Pop, Lock, Crane, Samba, Handstand and Kickup.

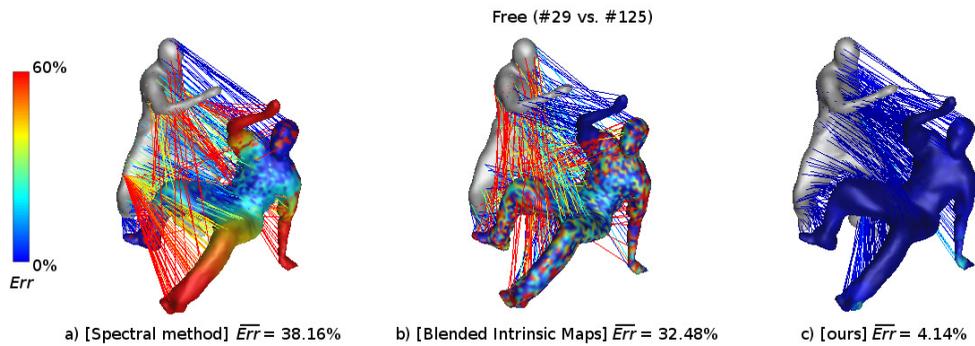


Fig. 12. Wide-timeframe matching. Alignment errors (distance to ground truth) between frames (#29 vs. #125) of sequence Free: a) [32] returns $Err = 38.16\%$, b) [31] return $Err = 32.48\%$, and c) [ours] returns $Err = 4.14\%$. (10% of matchings are represented.)

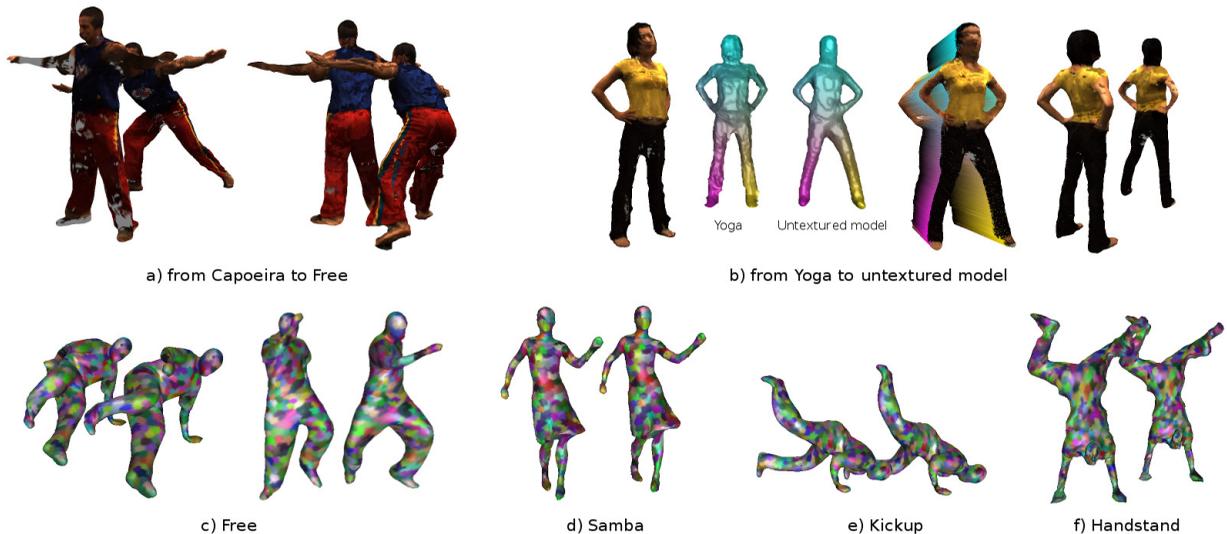


Fig. 13. Texture transfer by geodesic mapping. a) Texture from model of sequence Capoeira transferred to model of sequence Free. b) Texture from model of sequence Yoga transferred to non-textured model. c) d), e) and f) Textures from various models of [6] transferred for ground-truth evaluation (models on the left to models on the right). Mappings are very accurate, even between high-resolution models from different 3D video sequences.

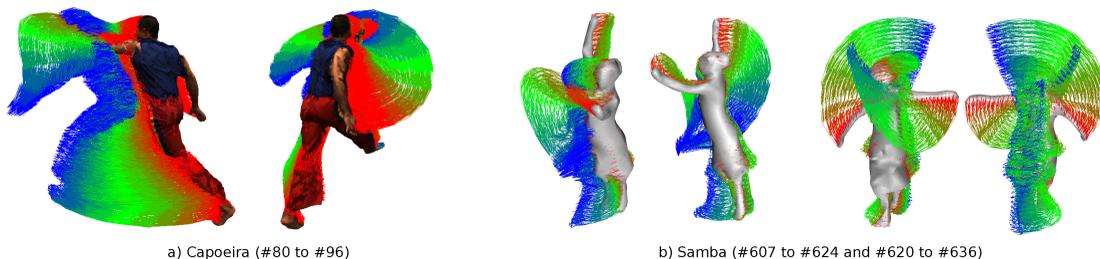


Fig. 14. 3D video motion flows (colors represent displacement sizes). a) Frames #80 to #96 from sequence Capoeira. b) Frames #607 to #624 and frames #620 to #636 from sequence Samba: motion flows are extracted even in poorly textured regions.

- [13] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance capture from sparse multi-view video," *ACM Trans. Graphics*, vol. 27, no. 3, 2008.
- [14] Y. Furukawa and J. Ponce, "Dense 3D motion capture from synchronized video streams," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [15] M. Liao, Q. Zhang, H. Wang, R. Yang, and M. Gong, "Modeling deformable objects from a single depth camera," *Proc.*

IEEE Int'l Conf. Computer Vision, 2009.

- [16] H. Jiang, H. Liu, P. Tan, G. Zhang, and H. Bao, "3D reconstruction of dynamic scenes with multiple handheld cameras," *Proc. European Conf. Computer Vision*, 2012.
- [17] F. Lazarus and A. Verroust, "Three-dimensional metamorphosis: a survey," *The Visual Computer*, vol. 14, pp. 373–389, 1998.
- [18] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow," *IEEE Trans. Pattern Analysis*

- Machine Intelligence*, vol. 27, no. 1, pp. 475–480, 2005.
- [19] J. Shi and C. Tomasi, “Good features to track,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1994.
- [20] E. de Aguiar, C. Theobalt, C. Stoll, and H.-P. Seidel, “Markerless deformable mesh tracking for human shape and motion capture,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
- [21] O. Sorkine, “Differential representations for mesh processing,” *Computer Graphics Forum*, vol. 25, no. 4, 2006.
- [22] H. Bay, T. Tuytelaars, and L. van Gool, “Surf: Speeded up robust features,” *Proc. European Conf. Computer Vision*, 2006.
- [23] J. Starck and A. Hilton, “Spherical matching for temporal correspondence of non-rigid surfaces,” *Proc. IEEE Int’l Conf. Computer Vision*, 2005.
- [24] N. Ahmed, C. Theobalt, C. Ross, S. Thrun, and H.-P. Seidel, “Dense correspondence finding for parametrization-free animation reconstruction from video,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [25] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int’l Journal Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [26] M. Wardetzky, S. Mathur, F. Kalberer, and E. Grinspun, “Discrete Laplace operators: No free lunch,” *Eurographics Symp. on Geometry Processing*, 2007.
- [27] A. Tevs, M. Bokeloh, M. Wand, A. Schilling, and H.-P. Seidel, “Isometric registration of ambiguous and partial data,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.
- [28] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Calculus of non-rigid surfaces for geometry and texture manipulation,” *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 5, 2007.
- [29] C. Stoll, Z. Karni, and H.-P. Seidel, “Geodesics guided constrained texture deformation,” *Pacific Graphics*, 2007.
- [30] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot, “Gromov-Hausdorff stable signatures for shapes using persistence,” *Computer Graphics Forum*, vol. 28, no. 5, pp. 1393–1403, 2009.
- [31] V. G. Kim, Y. Lipman, and T. Funkhouser, “Blended intrinsic maps,” *SIGGRAPH*, 2011.
- [32] H. Lombaert, L. Grady, J. R. Polimeni, and F. Cheriet, “Focus: Feature oriented correspondence using spectral regularization - a method for accurate surface matching,” *IEEE Trans. Pattern Analysis Machine Intelligence*, 2013.
- [33] T. Tung and T. Matsuyama, “Dynamic surface matching by geodesic mapping for 3D animation transfer,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [34] Y. Wang, B. S. Peterson, and L. H. Staib, “Shape-based 3D surface correspondence using geodesics and local geometry,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000.
- [35] Q.-X. Huang, B. Adams, M. Wicke, and L. J. Guibas, “Non-rigid registration under isometric deformations,” *Computer Graphics Forum*, vol. 25, no. 5, pp. 1449–1457, 2008.
- [36] H. Li, B. Adams, L. J. Guibas, and M. Pauly, “Robust single-view geometry and motion reconstruction,” *ACM Trans. on Graphics (Proc. SIGGRAPH Asia 2009)*, vol. 28, no. 5, 2009.
- [37] H. Li, L. Luo, D. Vlastic, P. Peers, J. Popović, M. Pauly, and S. Rusinkiewicz, “Temporally coherent completion of dynamic shapes,” *ACM Trans. on Graphics*, vol. 31, no. 1, 2012.
- [38] C. Cagniard, E. Boyer, and S. Ilic, “Iterative mesh deformation for dense surface tracking,” *Proc. IEEE Int’l Workshop on 3-D Digital Imaging and Modeling*, 2009.
- [39] R. W. Sumner and J. Popovic, “Deformation transfer for triangle meshes,” *SIGGRAPH*, pp. 399–405, 2004.
- [40] T. Tung and F. Schmitt, “The augmented multiresolution Reeb graph approach for content-based retrieval of 3D shapes,” *Int. Jour. of Shape Modeling*, vol. 11, no. 1, pp. 91–120, 2005.
- [41] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, “Topology matching for fully automatic similarity estimation of 3D shapes,” *SIGGRAPH*, pp. 203–212, 2001.
- [42] M. Morse, “The calculus of variations in the large,” *American Mathematical Society, Colloquium Publication 18*, New York, 1934.
- [43] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas, “Robust on-line computation of Reeb graphs: Simplicity and speed,” *SIGGRAPH*, 2007.
- [44] P. Huang, A. Hilton, and J. Starck, “Shape similarity for 3D video sequences of people,” *Int’l J. Computer Vision (IJCV)*, vol. 89, no. 2-3, pp. 362–381, 2010.
- [45] P. Huang, T. Tung, S. Nobuhara, A. Hilton, and T. Matsuyama, “Comparison of skeleton and non-skeleton shape descriptors for 3D video,” *Proc. Int’l Symp. 3D Data Processing, Visualization, and Transmission (3DPVT)*, 2010.
- [46] H. Edelsbrunner, J. Harer, A. Mascarenhas, and V. Pascucci, “Time-varying Reeb graphs for continuous space-time data,” *Symp. Computational Geometry*, 2004.
- [47] T. Klein and T. Ertl, “Scale-space tracking of critical points in 3D vector fields,” *Proc. Topology-Based Methods in Visualization*, 2005.
- [48] D. Vlastic, I. Baran, W. Matusik, and J. Popovic, “Articulated mesh animation from multi-view silhouettes,” *ACM Trans. Graphics*, vol. 27, no. 3, 2008.
- [49] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [50] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, “A comparative study of energy minimization methods for Markov random fields with smoothness-based priors export,” *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.
- [51] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Topology-invariant similarity of nonrigid shapes,” *Int’l Jour. Computer Vision (IJCV)*, vol. 81, pp. 281–301, 2009.
- [52] J. Starck and A. Hilton, “Correspondence labelling for wide-timeframe free-form surface matching,” *Proc. IEEE Int’l Conf. Computer Vision*, 2007.



Tony Tung received the M.Sc. degree in Physics and Computer Science from Télécom Physique, France, with a double degree in Photonics and Image Processing in 2000, and the Ph.D. degree in Signal and Image processing from Télécom ParisTech, France, in 2005. He worked as IT consultant (2000-2002) and senior R&D engineer (2005-2008) in private companies, and as postdoctoral research fellow at Kyoto University (2005, 2008-2009). Since 2010, he is an Assistant Professor at Kyoto University, working jointly at the Graduate School of Informatics, and at the Academic Center for Computing and Media Studies. His research interests include computer vision, pattern recognition, shape modeling, and human-machine interaction. He was awarded Fellowships from the Japan Society for the Promotion of Science in 2005 and 2008, and Grant-in-Aid for Young Scientists in 2011. He is a member of the IEEE.



Takashi Matsuyama received B. Eng., M. Eng., and D. Eng. degrees in electrical engineering from Kyoto University, Japan, in 1974, 1976, and 1980, respectively. He is currently a professor in the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. His research interests include knowledge-based image understanding, computer vision, 3D video, human-computer interaction, and smart energy management. He wrote more than 100 papers and books including two research monographs, *A Structural Analysis of Complex Aerial Photographs*, PLENUM, 1980 and *SIGMA: A Knowledge-Based Aerial Image Understanding System*, PLENUM, 1990. He won ten best paper awards from Japanese and international academic societies including the Marr Prize at ICCV’95. He is on the editorial board of the *Pattern Recognition Journal*. He was awarded Fellowships from the International Association for Pattern Recognition, the Information Processing Society of Japan, and the Institute of Electronics, Information, and Communication Engineers Japan. He is a member of the IEEE.