

Dynamic Surface Matching by Geodesic Mapping for 3D Animation Transfer

Tony Tung Takashi Matsuyama
Graduate School of Informatics, Kyoto University, Japan
{tung, tm}@vision.kuee.kyoto-u.ac.jp

Abstract

This paper presents a novel approach that achieves complete matching of 3D dynamic surfaces. Surfaces are captured from multi-view video data and represented by sequences of 3D manifold meshes in motion (3D videos). We propose to perform dense surface matching between 3D video frames using geodesic diffeomorphisms. Our algorithm uses a coarse-to-fine strategy to derive a robust correspondence map, then a probabilistic formulation is coupled with a voting scheme in order to obtain local unicity of matching candidates and a smooth mapping. The significant advantage of the proposed technique compared to existing approaches is that it does not rely on a color-based feature extraction process. Hence, our method does not lose accuracy in poorly textured regions and is not bounded to be used on video sequences of a unique subject. Therefore our complete surface mapping can be applied to: (1) texture transfer between surface models extracted from different sequences, (2) dense motion flow estimation in 3D video, and (3) motion transfer from a 3D video to an unanimated 3D model. Experiments are performed on challenging publicly available real-world datasets and show compelling results.

1. Introduction

Dynamic surface matching and tracking have become a growing research field due to the recent progresses in multi-view video reconstruction (3D video). 3D video consists in a free-viewpoint video of real-world subjects in motion immersed in a virtual world. In order to perform the acquisition, several calibrated and synchronized video cameras are set around the scene (e.g. a studio or a stadium). Subjects are then captured without wearing any special markers as opposed to motion capture methods (mocap). Thus, the technique suits to a very wide range of applications such as surveillance, medicine, sports, entertainment, etc.

3D video reconstruction returns a stream of textured surface models undergoing free-form deformation and represented by 3D manifold meshes. As every 3D video frame

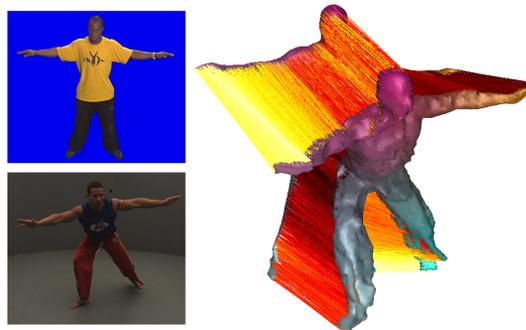


Figure 1. Complete surface matching of models from different sequences by geodesic mapping.

is reconstructed independently, surface matching and tracking problems are not trivial. To date, most of the existing approaches rely on color-based feature extraction to derive correspondences between frames. As a consequence, tracking can be lost in poorly photo-consistent regions (due to lack of texture or uncalibrated colors of image cameras), and matching is still limited to sequences of a unique textured model.

We present a novel technique that achieves complete mapping of dynamic surfaces using geodesic diffeomorphisms. The proposed method performs matching and tracking using shape characteristics rather than appearance features (such as colors or corners). Hence the algorithm is invariant to photometric inconsistencies and can handle 3D video sequences of different models (cf. Fig. 1). Our approach involves a coarse-to-fine mapping strategy to derive a robust correspondence map, and a voting scheme in a probabilistic formulation to obtain local unicity of matching candidates. The complete surface mapping enables us to perform: (1) texture transfer between two models from different sequences, (2) dense motion flow extraction in 3D video, and (3) motion transfer from a 3D video to an unanimated 3D model.

The rest of the paper is organized as follows. The next section discusses work related to the techniques presented in this paper. Section 3 presents our geodesic mapping strat-

egy. Section 4 describes the 3D surface matching process. Section 5 shows experimental results. Section 6 concludes with a discussion on our contributions.

2. Related work

An increasing number of multi-view video capture systems were developed in recent years [13, 16, 10, 7, 2, 21, 8, 11] (cf. Fig. 2). In the same way, the 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 our framework, our goal is to achieve complete surface matching of 3D video frames for texture and motion transfer.

The literature has provided several methods directly related to dense surface matching and tracking in 3D video. In general, the process consists in: (1) finding a sparse set of correspondences between frames and (2) achieve dense matching by regularization method or mesh deformation. The correspondence problem is indeed particularly well-known for surface mesh edition and morphing (cf. [14] for a survey). Nevertheless it is often required that the user interactively defines numerous pairs of markers on the models to match, which is impracticable with long 3D video sequences. Hence, techniques for automatic surface feature extraction and matching are still actively under investigation.

In [28], 3D scene flows are presented as an extension of optical flows [17] to represent the 3D motion of points. Optical flows are extracted and regularized in images and then used to derive the scene flows. In [9], scene flows and Laplacian mesh editing [18] are combined to track the deformation of a high-quality a priori shape model obtained by laser-scan. In [27], the problem is addressed by combining Speeded Up Robust Features (SURF) extraction [3] 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 [19], temporal correspondences are estimated by spherical parameterization and matching. This approach suits only to surfaces of genus-0. In [21], 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 [11], meshes are deformed over time by tracking photo-consistent sur-

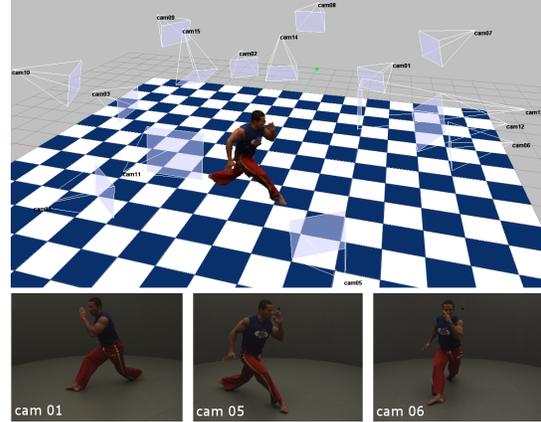


Figure 2. **3D video reconstruction from 15 video cameras and corresponding video frames.**

face patches and optimization processes. In [1], SIFT [15] are used as 2D features to obtain sparse correspondences between adjacent frames. Thereafter, dense correspondences are generated by mapping harmonic functions [29] 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 [30], 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 [24], 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. On the other hand, photometric feature matching approaches require surface models with good texture and color consistency between the multi-view cameras, which is in practice tedious to manage. Moreover high resolution cameras are necessary to achieve dense matching. Besides, using photometric features bounds the mesh matching to sequences of a unique textured model. In [5], the authors use geodesics to establish dense correspondence assuming isometry. Their approach uses a generalized multidimensional scaling (GMDS) which relies on an iterative minimization algorithm. The GMDS technique suits to low resolution models (100 vertices) but in our framework models contain up to 140,000 vertices.

In this paper, we present a novel approach for dense matching of dynamic surfaces in 3D videos using geodesic maps. Our technique does not involve any photometric features, thus it can be applied to match surface meshes from different sequences. As the cyclic nature of harmonic func-

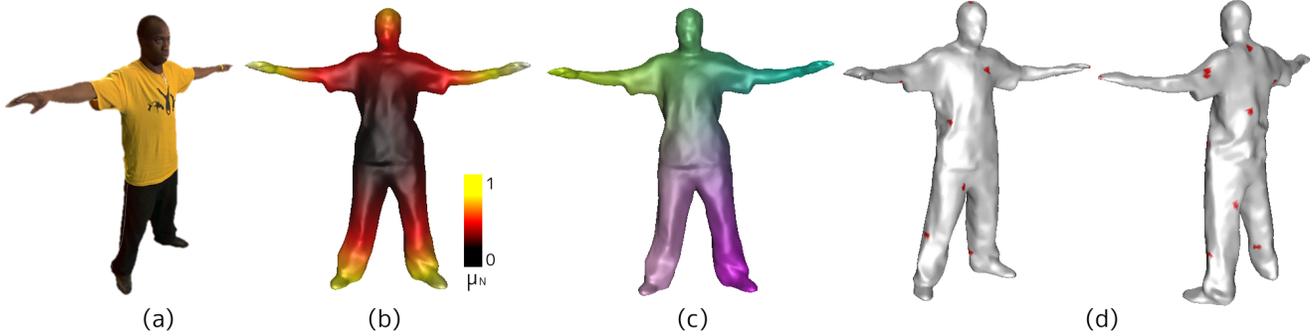


Figure 3. **Reference point extraction.** (a) Input mesh. (b) Extremal points are obtained from the geodesic integral values. (c) Geodesic distances to reference points are used to locate surface points. (d) Additional reference points improve the geodesic coordinate accuracy.

tions can cause mappings not to be bijective, we use vector fields of geodesic distances to perform a dense mapping between any surface points [22]. Extremal points of a geodesic integral function [12] defined on the surface model serve as reference points from which a sparse set of correspondences can be derived. Thus the method can be considered as a parameterization constrained by the correspondence points. The problem has been formulated as an MRF energy to minimize with smoothness assumption [23]. Our approach returns complete surface matchings of mesh models from different sequences. As a consequence it can be applied for texture and motion transfer, as well as dense 3D motion flow estimation.

3. Geodesic distance mapping

We propose to use geodesic diffeomorphisms as an alternative to local photometric features to determine correspondences between surfaces. Texture in real-world 3D video data is indeed not always consistent from multiple views due to lighting conditions. Hence surface tracking can become challenging. Moreover using geometric features enables us to perform matching between models of different sequences. The following subsections present our strategy to determine correspondences between two surface meshes. First, robust reference points are extracted using the global shapes of the surface models. This enables us to uniquely localize any surface point using barycentric coordinates. Thereafter in order to reinforce local accuracy, an additional set of reference points are obtained from a random selection over the surface. Thus we compute vector fields containing normalized geodesic distances to all reference points. Finally, the correspondence problem between two surfaces is solved using an energy minimization problem embedded in a coarse-to-fine strategy.

3.1. Robust reference point extraction

We assume 3D surfaces approximated by compact 2-manifold meshes. Let $S = (V, E)$ be a surface where V is a finite set of vertex positions in \mathbb{R}^3 and E defines the edge connections. According to the Morse theory, a continuous function μ defined on S characterizes the topology of the surface on its critical points. We choose the integral geodesic as Morse function μ , as defined in [12]:

$$\mu(\mathbf{v}) = \int_{\mathbf{p} \in S} g(\mathbf{v}, \mathbf{p}) dS \quad \text{and} \quad \mu_N(\mathbf{v}) = \frac{\mu(\mathbf{v}) - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \quad (1)$$

where $g(\mathbf{v}, \mathbf{p})$ is the geodesic distance on S between two points \mathbf{v} and \mathbf{p} belonging to S , and $\mu_N : S \rightarrow [0, 1]$ is the function μ normalized with respect to its minimal and maximal values μ_{\min} and μ_{\max} . μ_N is then invariant to rotation, translation and scale transformations. Moreover the integral formulation provides robustness to local surface noise such as outliers (e.g. due to 3D reconstruction artifacts). Extremal values of μ_N return surface critical point locations which coincide to highly concave or convex regions. Let $f : S_1 \rightarrow S_2$ be a mapping (diffeomorphism) between two surfaces $S_1 = (V_1, E_1)$ and $S_2 = (V_2, E_2)$. The points $p \in S_1$ and $q \in S_2$ are matched by computing $\min(\mu(p), \mu(q)) + \min(a(p), a(q))$ where $a(p)$ is the relative surface area around p having $\mu(p)$ values [12, 25].

In our experiments on real-world datasets, 3 to 5 maximal points corresponding to the tip of head and limbs are usually found. As three points are sufficient to define a unique barycentric coordinate system, every vertex can be uniquely identified by geodesic coordinates (a vector field containing geodesic distances to reference points). This statement is valid if and only if the three points are non-collinear which is in practice always true with real-world data. We obtain the correspondence map f between two surface meshes where the r_1 matching maximal values serve as robust reference points (constraints) (cf. Fig. 3).

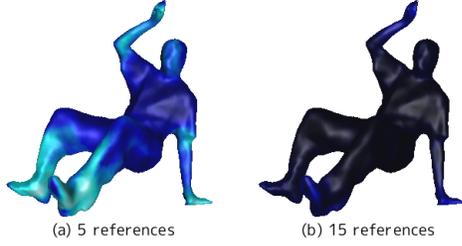


Figure 4. **Mapping Accuracy.** The surface colors indicate the *uniqueness* degree of each vertex with respect to (a) 5 reference points and (b) 15 reference points sampled on the surface. Light colors indicate regions of potential inaccuracies and dark colors indicate high confidence area.

3.2. Additional reference points

In 3D video the dynamic surface models are undergoing free-form deformations which can affect geodesics between two arbitrary frames (e.g. the deformations are not always isometric with large cloths). Hence, to guarantee geodesic consistency we introduce a set of r_2 additional reference points $\{R_i\} \in V_1$. The r_2 points are randomly chosen on the surface using an iterative process and are constraint to be at least distant to each other by τ to the previously selected references. The threshold τ is set to obtain an homogeneous coverage of the whole surface with a sufficient density. Every point p of the surface is then identified by a vector field of size $r = r_1 + r_2$ containing geodesic distances to all the reference points $v(p) = [\mu_1(p), \dots, \mu_r(p)]$. Let R_i and T_j be two reference points. If (R_i, T_j) is a matching pair, then:

$$\forall (R_i, T_j) \in (S_1, S_2), D_g(R_i, T_j) < \tau, \quad (2)$$

$$\text{where } D_g = \sum_{k=1}^r \|\mu_k(R_i) - \nu_k(T_j)\|_2, \quad (3)$$

D_g defines the distance between two geodesic coordinates, and μ_k and ν_k are the geodesic distances to the k^{th} reference points on S_1 and S_2 respectively. Note that μ_k and ν_k are normalized to their maximal values in order to be invariant to the model scale. A few set of reference points is practically enough to uniquely localize any surface point. Nevertheless geodesics which are too distant to reference points may contain errors due to surface noise or strong elastic deformation. The additional reference points guarantee that each surface point can be uniquely identified by at least n correct reference points using geodesic coordinates, while the remaining coordinates are possibly inaccurate (cf. Fig. 4). As a matter of fact, each reference point introduces a local isometric mapping, and having reference points close to regions where topology changes occur reduce matching errors (cf. Fig. 5). Hence in practice D_g should only take into account the n closest reference points.

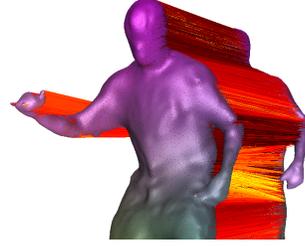


Figure 5. Matching artifacts due to change in surface topology can be compensated if some reference points lie in the area.

3.3. Correspondence matching

In our framework the correspondence problem consists in matching the reference points $\{R_i\} \in V_1$ to a set of points $\{T_j\} \in V_2$, such that (R_i, T_j) forms a pair if and only if:

$$T_j = \arg \min_{T_k \in V_2} (D_g(R_i, T_k)) \quad \text{with } D_g(R_i, T_j) < \tau_2, \quad (4)$$

where the threshold τ_2 prevents the matching of points that are too distant to the reference points. In addition, we set a voting system to guarantee the correctness of the matching, such that (R_i, T_j) is a valid pair if and only if $\exists m \geq n$,

$$\forall k \in [1, m], |\mu_k(R_i) - \nu_k(T_j)| < \tau_3, \quad (5)$$

where in practice $\tau_3 > \tau_2$, and $n = 8$. As the number of reference points r_2 is in the order of 20-25, a greedy search over all the entire mesh is not optimal, in particular for high resolution meshes (e.g. $|V_2| \sim 40.000$ candidates). We propose to adopt a coarse-to-fine strategy which embeds the vote count to reduce the problem complexity. By experiments, we observed that we could still obtain an accurate estimation of the matching points $\{T_j\}$ by starting the search on a reasonable low resolution mesh (e.g. 3000 vertices, 10000 edges). For each geodesic coordinates $v(T_j^c)$ obtained on the coarse mesh, we locally extract the closest adjacent vertices $\mathcal{N}(T_j^c)$ in the fine mesh (e.g. 3-ring neighbors). The geodesic coordinates $v(T_j)$ on the fine mesh are then obtained by local optimization of $v(T_j^c)$ in $\mathcal{N}(T_j^c)$. Therefore, we express the maximum a posteriori Markov Random Field (MAP-MRF) problem where the posterior probability to maximize is:

$$p(T_j | \mathcal{N}(T_j^c)) \propto \prod_j D_g(T_j, T_j^c) \prod_j \prod_{T_k \in \mathcal{N}(T_j)} V(T_j, T_k), \quad (6)$$

$$\text{where } V(T_j, T_k) = |D_g(T_j, T_k) - D_g(T_j^c, T_k^c)|, \quad (7)$$

and V is the smoothness assumption. A Loopy Belief Propagation method can be used to solve this problem [23]. Finally, we obtain the correspondence map f between two surface meshes with an extended set of matching reference

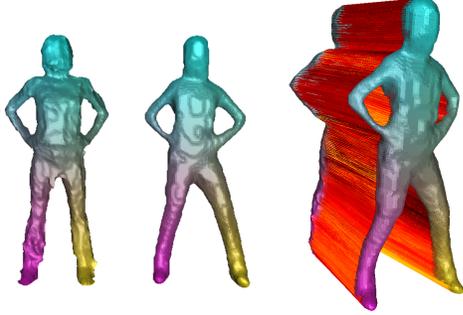


Figure 6. **Dense matching between two meshes from different 3D video sequences.** Every point on the destination mesh is mapped to a point on the source mesh. The colors encode geodesic distances to 3 reference points located on the head and at the feet of the models.

points on each surface (cf. Fig. 3). In 3D video, frames are reconstructed independently. Hence surface meshes do not share the same connectivity and $|V_1| \neq |V_2|$ (although usually similar for consecutive frames of a same sequence). In practice, it is therefore impossible to define f as a true bijection without inserting extra vertices on both surfaces. In our implementation, the morphism returns nearest neighbor approximations on the destination mesh. However as we deal with quite high resolution meshes, no artifact are visible.

Furthermore, in the case of consecutive frames from a same sequence, we can directly compute the probability $p(T_t)$ of a candidate $T_t \in V_2$ to match the reference $R_t \in V_1$ at t . According to Bayes’ theorem, the posterior probability $p(T_{t+1})$ at $t + 1$ can be estimated using $p(T_t)$ as prior probability:

$$p(T_{t+1}) = \frac{p(T_t)E_R(T_t)}{p(T_t)E_R(T_t) + (1 - p(T_t))}, \quad (8)$$

where $E_R(T_t) = \Gamma(D_g(T_t, R_t))$, Γ being a Gaussian distribution. Consequently, the probability for a candidate T_{t+1} to match the reference R_t is estimated to $p(T_{t+1})$ at $t + 1$ and can be used to reduce search spaces.

4. Surface matching

We present an MRF energy minimization formulation to achieve dense matching of 3D surfaces while keeping rigidity constraints between the matching vertices, and where the matching references serve as priors.

4.1. Energy formulation

Let $\mathcal{P} = \{p_i\}$ be the set of vertex sites on a surface mesh S_1 , and $\mathcal{L} = \{l_p\}$ be a discrete set of labels corresponding to the candidate vertices on a surface mesh S_2 . We propose

to minimize the following MRF energy in order to achieve the complete mapping of S_1 onto S_2 :

$$E(f) = E_d(f) + E_s(f). \quad (9)$$

$E(f)$ is the 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:

$$E_d(f) = \sum_{p \in \mathcal{P}} D_g(p, l_p), \quad (10)$$

$$E_s(f) = \sum_{\{p,q\} \in \mathcal{N}} \lambda \cdot |T_p(l_p) - T_q(l_q)|, \quad (11)$$

where \mathcal{N} is a neighborhood configuration in S_2 , λ is a constant factor, and $T_p(l_p)$ and $T_q(l_q)$ are *vector flows* between p and l_p , and between q and l_q respectively. This is an NP-complete problem which minimization is indeed very computationally expensive using the usual techniques [4, 23].

However in our framework we can dramatically reduce the label search space by splitting the global optimization problem into smaller local ones, and consequently speed up the optimization process. This is allowed as the problem constraints are independent of each other. The strategy consists in sweeping the surface mesh using front propagations from reference points in order to expand their matchings to neighbor regions. As a site vertex $p \in S_1$ is visited, we can determine the corresponding search space on S_2 containing the matching vertex l_p by taking advantage of prior matches on sites $\{q\} \subset \mathcal{N}(p)$, where $\mathcal{N}(p)$ is the neighborhood of p (e.g. the 3-ring adjacent vertices). Our assumption is that if $q \in \mathcal{N}(p)$, then $l_p \in \mathcal{N}(l_q)$. Thereafter l_p can be determined by minimizing the energy in $\bigcup_{\{q\} \subset \mathcal{N}(p)} \mathcal{N}(l_q)$, as defined in Eq. 9. Figure 6 illustrates a dense matching between two models of different sequences. The color codes represent the distances to 3 reference points located on the head and at the feet of the models, although 15 reference points were used to achieve the complete mapping. Note that the models have consequent noise on the surface, making the use of local geometric descriptors (e.g. [30]) impossible to use in this case.

5. Experimental results

The algorithms were developed in C++ using a PC Core2Duo 3.0GHz 4GB RAM. For our evaluation, we use publicly available academic databases of 3D video sequences, such as the *free* sequence from the University of Surrey [21], the *dance* model from INRIA Grenoble [2], and the *yoga* and *capoeira* sequences from Kyoto University [16]. In the *free* sequence, the subject wears a loose T-shirt and performs quick dance steps. Every mesh contains ~ 140.000 vertices and has a very smooth surface. The *dance* sequence consists in a dancer performance wearing a gown. Every mesh has low resolution and contains ~ 1.000

vertices. Both the *yoga* and *capoeira* sequences contain high resolution meshes of $\sim 15,000$ vertices, but with a noisy surface (with bumps). The computation time between two surface is proportional to the size of both source and destination meshes ($\sim 20s$ for 30,000 vertices). The most time consuming step in the computation of the geodesic integral (cf. Sect. 3.1). We implemented the Dijkstra algorithm to compute shortest paths on mesh surface. The computation complexity is $\mathcal{O}(N \log N)$, N being the number of vertices.

Our algorithm is able to robustly handle different kinds of applications such as texture transfer, wide-timeframe matching [20], 3D motion flow estimation [27], and motion transfer [6]. In order to evaluate the accuracy of our matching, we have computed the flows between consecutive frames and displaced the source mesh vertices following the flows. Then we could observe the difference between the displaced mesh and the destination mesh. We have indeed observed very minor differences which were mainly due to the triangulation difference. In the following subsections, the results are compared to several state-of-the-art techniques.

5.1. Texture transfer

As illustrated in Fig. 8, the accuracy of our approach allows us to perform texture transfer between surfaces. The *yoga* performer is matched with an untextured surface model, and transfers her texture to it (left), and the *capoeira* performer transfers his texture to the model from the *free* sequence (right). (Note that the previously untextured model appears in Fig. 6 with the *yoga* model without texture as well.) The texture transfer returns globally a good visual result although some local and minor artifacts can be observed. In particular the wrinkles on the shirts are well mapped. Obviously the texture transfer performance is optimal (i.e. without distortion) if and only if the source and destination are identical (up to a global scale factor, as the geodesic coordinates are normalized).

5.2. Wide-timeframe matching

We perform dense surface matchings between wide-timeframe surface meshes. The process can be achieved without a prior model or frame-to-frame surface tracking, as [20]. However, our approach return a complete matching of the surface, where any point is mapped. The Figure 9 illustrates the wide-timeframe surface mapping on meshes from the *free* sequence. The frame #0 is matched with the frame #125 (left), and with the frame #29 (right). We believe our mapping outperforms [20] in terms of density.

5.3. 3D Motion flow

Motion flow is obtained by matching the consecutive frames of a video sequence and allows to visually appreciate the different trajectories of the vertices. The Figure 10 show the motion flow computed on the *dance* sequence. Our results can be somehow compared to the one presented in [27]. Nevertheless as we do not deform a mesh over time, we do not have the connectivity consistency between consecutive frames, and no vertex tracking over time is performed. In our framework, the flows are computed by considering successive pairs of frame and every vertex is mapped. Thus we obtain dense motion fields. The Figure 11 show the motion flow computed on the *capoeira* sequence from two different views. Again, we observe dense motion fields.

5.4. Motion transfer

Motion transfer is performed by applying the trajectories of the vertices of a source mesh to the vertices of a destination mesh. This can be naturally achieved using our approach as our matching scheme always delivers a complete mapping between consecutive frames (cf. Fig 7). In practice a Laplacian editing method is applied to displace the vertices in order to smooth out outliers [18]. However, the source and destination meshes should possess a certain degree of similarity, otherwise the motion transfer would not make sense. In order to extract similar meshes for 3D video sequences, we can use a topology descriptor such as in [26], where a topology-based similarity measure is used to retrieve and match the models belonging to the same topology class.

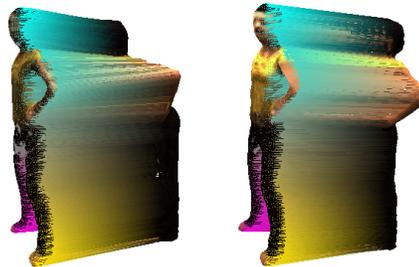


Figure 7. **Motion transfert between mesh models.** Unanimated models can be easily deformed under the guidance of a dense extracted motion flows.

6. Conclusion

In this paper, we present a novel approach to perform dense matching of 3D dynamic surfaces captured by a set of multi-view video camera (3D video). We propose to use geodesics to characterize the surfaces by deriving ref-

erence points which will serve to uniquely identify the surface points. The complete surface mapping is achieved using a probabilistic scheme which returns uniqueness and local smoothness of the solution. Our approach does not involve color-based feature extraction, and therefore is able to achieve matching of surface models from different sequences. Moreover it can be used for texture transfer, 3D motion flow estimation in 3D video, and motion transfer. The performances are obtained on publicly available real-world datasets. Further experiments to evaluate the accuracy of matching could be achieved by comparison with ground-truth manually marked correspondences.

Acknowledgments

This work was supported in part by the Japan Society for the Promotion of Science (JSPS No.20.08366), and JST-CREST project “Creation of Human-Harmonized Information Technology for Convivial Society”.

References

- [1] N. Ahmed, C. Theobalt, C. Rossl, S. Thrun, and H.-P. Seidel. Dense correspondence finding for parametrization-free animation reconstruction from video. *CVPR*, 2008. 2
- [2] J. Allard, C. Ménier, B. Raffin, E. Boyer, and F. Faure. Grimage: Markerless 3d interactions. *SIGGRAPH - Emerging Technologies*, 2007. 2, 5
- [3] H. Bay, T. Tuytelaars, and L. van Gool. Surf: Speeded up robust features. *ECCV*, 2006. 2
- [4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001. 5
- [5] 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*, 13(5), 2007. 2
- [6] G. K. M. Cheung, S. Baker, J. Hodgins, and T. Kanade. Markerless human motion transfer export. *3DPVT*, 2004. 6
- [7] K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking. *IJCV*, 63(3):225–245, 2005. 2
- [8] 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*, 27(3), 2008. 2
- [9] E. de Aguiar, C. Theobalt, C. Stoll, and H.-P. Seidel. Markerless deformable mesh tracking for human shape and motion capture. *CVPR*, 2007. 2
- [10] 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*, page 31, 2004. 2
- [11] Y. Furukawa and J. Ponce. Dense 3d motion capture from synchronized video streams. *CVPR*, 2008. 2
- [12] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii. Topology matching for fully automatic similarity estimation of 3d shapes. *SIGGRAPH*, pages 203–212, 2001. 3
- [13] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. A stereo machine for video-rate dense depth mapping and its new applications. *CVPR*, 1996. 2
- [14] F. Lazarus and A. Verroust. Three-dimensional metamorphosis: a survey. *The Visual Computer*, 14:373–389, 1998. 2
- [15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2
- [16] 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. *CVIU*, 96(3):393–434, 2004. 2, 5
- [17] J. Shi and C. Tomasi. Good features to track. *CVPR*, 1994. 2
- [18] O. Sorkine. Differential representations for mesh processing. *Computer Graphics Forum*, 25(4), 2006. 2, 6
- [19] J. Starck and A. Hilton. Spherical matching for temporal correspondence of non-rigid surfaces. *ICCV*, 2005. 2
- [20] J. Starck and A. Hilton. Correspondence labelling for wide-timeframe free-form surface matching. *ICCV*, 2007. 6
- [21] J. Starck and A. Hilton. Surface capture for performance-based animation. *IEEE Computer Graphics and Applications*, 2007. 2, 5
- [22] C. Stoll, Z. Karni, and H.-P. Seidel. Geodesics guided constrained texture deformation. *Pacific Graphics*, 2007. 3
- [23] 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. *PAMI*, 30(6):1068–1080, 2008. 3, 4, 5
- [24] A. Tevs, M. Bokeloh, M. Wand, A. Schilling, and H.-P. Seidel. Isometric registration of ambiguous and partial data. *CVPR*, 2009. 2
- [25] T. Tung and F. Schmitt. The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes. *Int. Jour. of Shape Modeling*, 11(1):91–120, 2005. 3
- [26] T. Tung and T. Matsuyama. Topology dictionary with markov model for 3d video content-based skimming and description. *CVPR*, 2009. 6
- [27] K. Varanasi, A. Zaharescu, E. Boyer, and R. Horaud. Temporal surface tracking using mesh evolution. *ECCV*, 2008. 2, 6
- [28] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. *PAMI*, 27(1):475 – 480, 2005. 2
- [29] M. Wardetzky, S. Mathur, F. Kalberer, and E. Grinspun. Discrete laplace operators: No free lunch. *Eurographics Symp. on Geometry Processing*, 2007. 2
- [30] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. *CVPR*, 2009. 2, 5



Figure 8. **Surface texture transfer.** Left: the *yoga* performer is matched with an non-textured surface model, and transfers her texture to it. Right: the *capoeira* performer transfers his texture to the model from the *free* sequence. We can observe a globally good mapping.

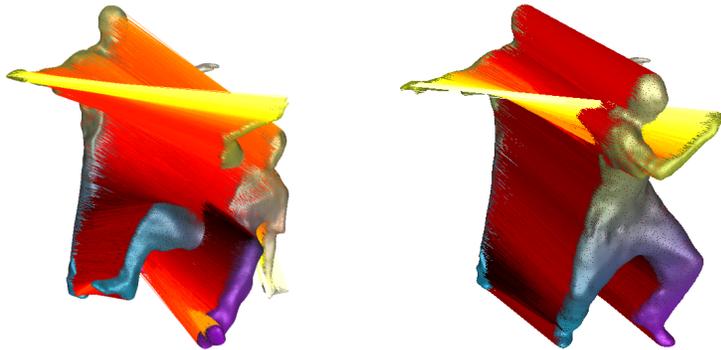


Figure 9. **Wide-timeframe matching.** Dense matching is performed between arbitrary frames of the *free* sequence. Left: frame #0 is matched with frame #125. Right: frame #0 is matched with frame #29. The links are color coded so that lighter colors (yellow) represent the most distant parts. We can observe a very dense matching.

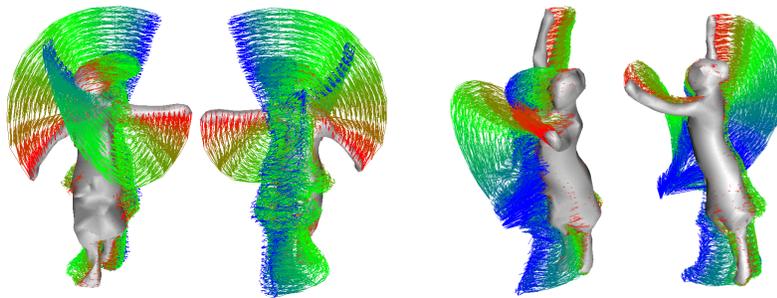


Figure 10. **Motion flow from the *dance* sequence.** Left: frames 620-636. Right: frames 607-624. Dense flows are extracted between consecutive frames.

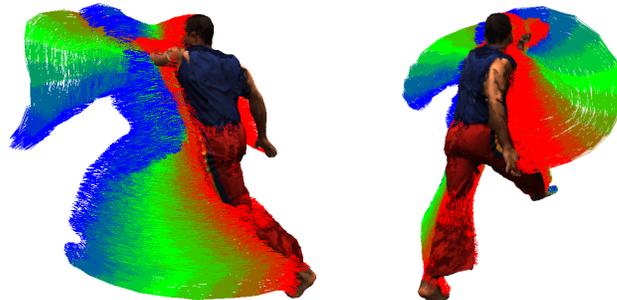


Figure 11. **Motion flow from the *capoeira* sequence.** Two views of frames 80-96. Dense flows are extracted between consecutive frames.