Please do not remove this page



UNIVERSITY OF
SURREY

Correspondence labelling for wide-timeframe free-form surface matching

Starck, J; Hilton, A

https://openresearch.surrey.ac.uk/esploro/outputs/conferencePresentation/Correspondence-labelling-for-wide-timeframe-free-form-surface/99515424002346/filesAndLinks?index=0

Starck, J., & Hilton, A. (2007). Correspondence labelling for wide-timeframe free-form surface matching. 1–8. https://doi.org/10.1109/ICCV.2007.4409108
Document Version: Text

Published Version: https://doi.org/10.1109/ICCV.2007.4409108

Surrey Open Research repository homepage: https://openresearch.surrey.ac.uk/esploro/
openresearch@surrey.ac.uk
SRIDA
Research:Open
Downloaded On 2026/03/15 05:01:56 +0000

Correspondence labelling for wide-timeframe free-form surface matching

Jonathan Starck and Adrian Hilton
Centre for Vision, Speech and Signal Processing,
University of Surrey, Guildford. GU2 7XH. UK
{ j.starck, a.hilton } @surrey.ac.uk

Abstract

This paper addresses the problem of estimating dense correspondence between arbitrary frames from captured sequences of shape and appearance for surfaces undergoing free-form deformation. Previous techniques require either a prior model, limiting the range of surface deformations, or frame-to-frame surface tracking which suffers from stabilisation problems over complete motion sequences and does not provide correspondence between sequences. The primary contribution of this paper is the introduction of a system for wide-timeframe surface matching without the requirement for a prior model or tracking. Deformation-invariant surface matching is formulated as a locally isometric mapping at a discrete set of surface points. A set of feature descriptors are presented that are invariant to isometric deformations and a novel MAP-MRF framework is presented to label sparse-to-dense surface correspondence, preserving the relative distribution of surface features while allowing for changes in surface topology. Performance is evaluated on challenging data from a moving person with loose clothing. Ground-truth feature correspondences are manually marked and the recall-accuracy characteristic is quantified in matching. Results demonstrate an improved performance compared to non-rigid point-pattern matching using robust matching and graph-matching using relaxation labelling, with successful matching achieved across wide variations in human body pose and surface topology.

1. Introduction

A framework is presented to match arbitrary frames from captured sequences of a free-form surface undergoing large scale articulated motions. The goal of this work is to perform surface matching in captured sequences of human motion as a basis for both intra-sequence and inter-sequence surface correspondence for the analysis of dynamic surface deformation. Intra-sequence correspondence requires matching between arbitrary frames in a single motion sequence and inter-sequence correspondence requires matching between arbitrary frames from two different sequences.

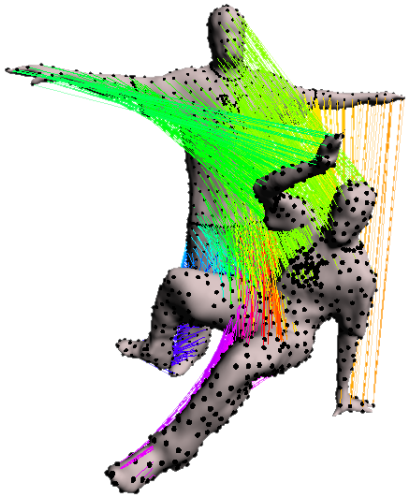


Figure 1. Correspondence labelling for wide-timeframe matching between two frames from different surface motion sequences.

Temporally varying surface matching has been addressed previously using model tracking and fitting, providing an approximate surface correspondence. Model based techniques however require a predefined model and restrict the space of feasible surface deformation [20]. In tracking, 3D *scene-flow* [23] has been proposed to derive temporal correspondence. However, stable long term flow is required for wide-timeframe matches which remains an open problem for differential tracking algorithms and temporal correspondence cannot match surfaces between different motion sequences. This paper contributes a new approach to the surface matching problem that allows recovery of surface correspondence in arbitrary time frames without the requirement for sequential tracking or a predefined model and with the flexibility to allow for changes in surface topology during motion.

Our approach to 3D surface matching is motivated by recent work in viewpoint invariant image matching for wide-baseline image correspondence [14]. Wide-baseline im-

age matching is addressed by constructing feature-point descriptors that are invariant to a specific class of transformations and by deriving a consistent set of feature matches that place the images in correspondence. Here, we formulate 3D surface correspondence as the problem of matching surface points between two objects using local descriptors that are invariant to surface deformation. Deformation is assumed to be a locally isometric transformation that preserves geodesic surface distance. Feature matching is then formulated as the inference of the most likely labelling of correspondence by modelling the feature point configurations as a Markov random field (MRF) that preserves the geodesic distance between feature points. A maximum a posteriori (MAP) estimate for the joint distribution of correspondence labels is then recovered by Bayesian belief propagation.

The primary contribution of this paper is a framework for wide-timeframe surface correspondence that does not require a prior model or temporal tracking. Two contributions are made in this framework, (i) a set of feature descriptors are presented that are invariant to isometric deformations and (ii) a novel MAP-MRF framework is presented to label sparse-to-dense surface correspondence. The result is a dense surface-to-surface correspondence that locally provides an isometric embedding while allowing for changes in global geodesic surface distance with changes in surface topology during deformation. An evaluation is performed in matching surface points across different poses in a public database of human surface motion [22]. The recall-accuracy characteristic, the number of correct matches against the required tolerance in matching is evaluated and compared to non-rigid point-pattern matching using robust matching (RPM-TPS) [4] and relaxation labelling [24].

2. Related Work

Conventional multiple-view scene reconstruction recovers a separate surface representation at each time instant and does not provide the structure and motion of the underlying scene. Structure and motion has been addressed using flow-based algorithms to track surfaces in multiple view video sequences. Vedula et al. [23] introduced the concept of *scene-flow* as the extension of 2D optic-flow to the 3D surface of a scene. Carceroni and Kutulakos [2] compute shape and non-rigid motion under known lighting conditions using relatively large scale 3D surface elements. Pons et al. [16] present a variational approach for reconstruction and scene-flow providing a global solution for the motion estimation of a surface. For cloth motion, distinct features have been used in tracking to produce a consistent surface parameterisation requiring either highly texture surfaces [17] or a colour-coded pattern [19]. Wide-timeframe motion-estimation is inherently ambiguous and an open problem for flow-based differential tracking algorithms. Correspondence has only been demonstrated for highly structured scenes [19] or by restricting the topology of the scene [21].

Surface correspondence has been addressed without the requirement for sequential tracking using prior models of shape and appearance. Model registration and fitting at arbitrary time points provides an approximate surface correspondence via the intermediate domain of the model surface. Salzmann et al. [18] construct a parameterised deformable model for the non-rigid motion of inextensible planar surfaces. Correspondence is derived in monocular image sequences using wide-baseline feature matching for registration and by fitting the parameterised model. In the medical domain, deformable models have been widely used to extract anatomical surfaces using landmark registration and model deformation [13]. For human surface shape, parameterised models are widely used to derive the gross articulated motion of the human body [15]. Deformable surface fitting has been performed [20, 3] given the pose of a skeleton model to extract the shape and appearance of a person at different time frames. Model-based matching and fitting is however inherently limited to the shape and topology of the prior parameterised model, which constrains the space of feasible surface deformations. For example, a generic human body will not match a person with loose clothing or dynamic hair motion.

Feature based surface matching has been investigated previously for the problem of appearance matching in images and shape matching for surfaces. In wide-baseline image matching [14] affine invariant feature descriptors are typically adopted as invariant to transformations between views. For monocular image sequences Ling and Jacobs [11] introduce a local appearance distribution descriptor for deformation-invariant image matching. Belongie et al. [1] proposed a local shape context descriptor for deformable 2D shape matching. Local feature descriptors have been developed for 3D shape matching in 3D shape recognition [9]. Gal et al. [6] introduce a local shape distribution descriptor that is invariant to articulated pose. Gatzke and Garland [7] resample surface curvature onto a radial descriptor embedded on the surface. Elad and Kimmel [5] construct bending invariant representations using an isometric embedding for a surface in a higher dimensional Euclidean space. In this paper we describe a set of bending invariant local descriptors for both shape and appearance embedded on a deforming surface.

3. Surface Descriptors

Surface matching is based on the assumption that deformations are inextensible, undergoing neither dilation or contraction during motion. The correspondence between two surfaces S and T can then be formalised as a map $f : S \rightarrow T$ that preserves geodesic surface distance, an isometric mapping. The strict assumption of a single distance-

preserving bijective map [5, 21] is relaxed. Large changes in geodesic distance can occur in practise with changes in surface topology for articulated motions and as surface regions appear and disappear with changes in visibility. Instead we assume only a local isometric mapping at a discrete set of surface points. In this section surface point descriptors are constructed that are invariant to isometric deformations and in Section 4 matching of point descriptors is formulated in a probabilistic framework that preserves geodesic surface distance where this is possible.

3.1. Isometric mapping

Local descriptors are constructed for a surface in an intermediate domain where the non-rigid isometric mapping between two surfaces can be represented by a rigid isometry. A local Euclidean isometric embedding is performed in $\mathbb{R}^2$, removing surface bending such that the mapping between two surfaces can be locally evaluated up to a 2D rigid body transformation. Isometric embedding has been proposed previously [5] using Multi-Dimensional Scaling (MDS) to preserve geodesic surface distance. Here Least-Squares Conformal Mapping (LSCM) [10] is adopted, minimising angular deformations as well as non-uniform surface scaling in constructing the map $f : S \rightarrow \mathbb{R}^2$.

A surface S is represented by a triangulated mesh defined by (V, K), where V is a finite set of vertex positions in $\mathbb{R}^3$ and K is a *simplicial complex* defining the vertices $\{i\} \in K$, edge connections $\{i, j\} \in K$, and face connections $\{i, j, k\} \in K$. A local mapping $f : S \rightarrow \mathbb{R}^2$ is centred at a given vertex i corresponding to a feature point on the surface. A genus-zero submesh is first extracted using a wavefront centred on i by extracting all triangles up to a given geodesic distance in which the triangle vertices have a consistent distance in the wavefront. This submesh is topologically equivalent to a disc and so can be flattened to $\mathbb{R}^2$ using LSCM [10]. The final step is to scale the map such that geodesic distance is consistent in $\mathbb{R}^2$ by applying an area preserving uniform scaling. Figure 2 illustrates the isometric-mapping for corresponding points on two surfaces.

3.2. Surface attributes

Surface attributes are resampled onto the geodesic-distance preserving domain for comparison. Attributes are defined for both the colour and shape at the vertices $\{i\} \in K$. Colour is derived by view-dependent blending from the multiple view images used to reconstruct the surface. A conservative visibility test is performed using a depth buffer in which the surface has been inflated such that ambiguous points close to occlusion boundaries become occluded and are not considered visible. Shape is represented using volumetric integral invariants providing robustness to surface noise compared to differential curvature based descriptors



Figure 2. Isometric mapping $f : S \rightarrow \mathbb{R}^2$ at a manually marked point correspondence between two surfaces. The appearance is similar up to a rigid isometry in $\mathbb{R}^2$.

[8]. While volumetric invariants in fact characterise surface bending and as such are not preserved in isometric mapping, they provide a means to distinguish dissimilar volumetric regions such as the human arms, legs and body. A single attribute vector is constructed for each vertex $\underline{a}_i$ with both colour and shape attributed, together with a covariance matrix Σ defining the expected variance for the attributes in motion sequences. The distance between attribute values can then be computed as the normalised sum of squared differences where N_a is the size of $\underline{a}$.

$$d\left(\underline{a}_i, \underline{a}_j\right) = \frac{1}{N_a}(\underline{a}_i - \underline{a}_j)^T \Sigma^{-1} (\underline{a}_i - \underline{a}_j) \qquad (1)$$

3.3. Corner descriptor

Corner features are widely used in matching and recognition tasks due to the localised and potentially distinct nature of the feature point. Corner features are extracted in the multiple-view images used to reconstruct a surface S using the SUSAN corner detector. The features are then associated with the closest visible vertex on the surface. A descriptor F_i^C is then constructed for each feature vertex by mapping to $\mathbb{R}^2$ and resampling attributes in spherical bins $F_i^C(r, \theta)$. The minimum distance between two feature descriptors is computed up to a rigid-body rotation of the descriptor.

$$d\left(F_i^C, F_j^C)\right) = \arg\min_{\phi} \frac{1}{N_{bins}} \sum_r \sum_\theta d\left(\underline{a}_i(r, \theta), \underline{a}_j(r, \theta + \phi)\right) \qquad (2)$$

3.4. Edge descriptor

Edge features define a locally salient region that is ambiguous in one orientation. Edge features are extracted using a Canny-Deriche edge detector in the multiple-view images. The features are again associated with the closest visible vertex on the surface and the vertex set is uniformly decimated to provide a semi-dense sampling along the sur-

face edges. Here conservative visibility testing serves to remove artificial edges that occur at occlusion boundaries in an image. A rotation invariant edge descriptor F_i^E is constructed by mapping to $\mathbb{R}^2$ and then resampling relative to a an assigned edge orientation as adopted in the SIFT feature descriptor [12]. The distance between two features is then computed with no transformation.

$$d\left(F_i^E, F_j^E\right)) = \frac{1}{N_{bins}} \sum_u \sum_v d\left(\underline{a}_i(u,v), \underline{a}_j(u,v)\right) \quad (3)$$

3.5. Region descriptor

Corners and edges provide only sparse feature points. Dense surface matching is performed by uniformly sampling the surface to create a set of region feature points that describe the relative distribution of surface attributes for non-salient regions. The *geodesic-intensity histogram* (GIH) [11] approach is adopted in which the distribution is represented by a 2D histogram of attribute values against geodesic distance. The descriptor provides a gross representation of the surface attributes that is invariant to distance preserving deformations. This descriptor was proposed to match intensity images and is applied here to a manifold in $\mathbb{R}^3$ with arbitrary surface attributes. The descriptor F_i^R is computed directly without mapping to $\mathbb{R}^2$ by traversing a wavefront centred on a feature vertex i and binning attribute values directly against geodesic distance. The distance between two descriptors is computed using the χ^2 difference between the normalised histograms.

$$d\left(F_i^R, F_j^R\right)) = \frac{1}{2N_a} \sum_a \sum_u \sum_v \frac{\left(F_i^R(u,v,a) - F_j^R(u,v,a)\right)^2}{F_i^R(u,v,a) + F_j^R(u,v,a)} \quad (4)$$

Figure 3 illustrates the corner, edge and region feature points extracted for a surface.

4. Correspondence Labelling

Surface correspondence is formulated as the problem of finding the most likely assignment between a set of feature points on a source surface S and target surface T. A local neighbourhood structure is defined on S forming an undirected graph, or Markov random field (MRF). The nodes of this graph represent random variables defining the discrete assignment from S to T and the edges define dependencies in the assignment for neighbouring points on S. The edge connections in the graph provide a prior probability on the compatibility between neighbouring point assignments and local support is aggregated in the graph through a process of belief propagation. The maximum a posterior (MAP) solution is sought in labelling the correspondence, in which the joint probability distribution for the assignment



Figure 3. Surface feature points showing extracted corners, edges and region points as black points on the surface mesh.

is maximised. In this section we outline the construction of a Markov network for the correspondence problem and present our algorithm for sparse-to-dense surface matching. Local support is explicitly defined in this formulation up to a given geodesic surface distance D_{max} for which topology and hence geodesic surface distance is assumed to be preserved, while allowing for changes in global surface topology at $d > D_{max}$.

4.1. MRF Formulation

The set of feature points on the source S is denoted by X where a feature point is defined by the position in $\mathbb{R}^3$ and the feature descriptor, $x_s = (\underline{x}_s, F_s) \in X$. Similarly the set of feature points on the target T is denoted by $y_t = (\underline{y}_t, F_t) \in Y$. The correspondence between feature points is a set of random variables L where $l_i \in L$ corresponds to each feature point $x_i \in X$ and takes one of a discrete set of labels corresponding to the feature points $y_t \in Y$. A particular realisation of the label variable l_i implies a specific correspondence between feature points $l_i : x_s = y_t$.

The likelihood that a source point corresponds to a particular target point is defined by a potential function $\Phi(l_i)$.

$$P\left(x_s = y_t | l_i\right) \propto \Phi\left(l_i\right) \quad (5)$$

A pair-wise Markov network is defined on S as an undirected graph $G = (N, E)$ where each node in the graph N represents the random variable l_i for assignment and each edge in E a dependency between two nodes l_i, l_j. The graph is constructed using a local neighbourhood system N_i that

contains all points within a given geodesic surface distance $D_N \leq D_{max}$ of the source feature point x_i.

$$N_i = \{j \in X | d(x_i, x_j) \leq D_N, j \neq i\} \quad (6)$$

The prior probability for an assignment l_i in the MRF is then defined by a compatibility function $\Psi(l_i, l_j)$ on the neighbourhood N_i. This compatibility function defines the probability of assignment according to consistency in labelling in the neighbourhood.

$$p(l_i) \propto \prod_{j \in N_i} \Psi(l_i, l_j) \quad (7)$$

The posterior probability at a node in the graph can then be written by applying Bayes' rule as follows, where Z is a normalising factor.

$$p(l_i|y_t) = \frac{1}{Z} \Phi(l_i) \prod_{j \in N_i} \Psi(l_i, l_j) \quad (8)$$

The MAP solution is found by maximising the joint posterior probability distribution.

$$L = \arg\max_{l} \prod_{i} \Phi(l_i) \prod_{i,j} \Psi(l_i, l_j) \quad (9)$$

The probability functions are modelled using a Gaussian distribution for the relative distance between feature points. The potential function $\Phi(l_i)$ defines the likelihood of a correspondence label according to the distance between the corresponding feature descriptors F_s, F_t.

$$\Phi(l_i) = \begin{cases} \exp\left\{-\lambda_F d(F_s^{SD}, F_t^{TD})\right\} & \text{SD = TD} \\ 0 & \text{SD} \neq \text{TD} \end{cases} \quad (10)$$

The compatibility function $\Psi(l_i, l_j)$ defines the prior probability for a realisation using the difference in the geodesic surface distance between the source points $d(x_i, x_j)$ and the labelled assignment $d(y_{l_i}, y_{l_j})$.

$$\Psi(l_i, l_j) = \exp\left\{-\lambda_D \left(d(x_i, x_j) - d(y_{l_i}, y_{l_j})\right)^2\right\} \quad (11)$$

The relative influence of the distance functions is controlled by a parameter λ. As we only compare features of the same type ($SD = TD$), there is no need to balance the relative weighting of feature distances and in practise we simply set $\lambda_F = 1$. The influence of the prior compatibility probability is set according to the expected variation in geodesic distance, a function of both differences in surface and feature sampling as well as non-isometric deformations arising from inexact surface reconstruction, for example where folds in clothing are not reconstructed leading to a reduced surface area.

4.2. Propagating Correspondence

Exact inference in the MRF is infeasible and the MAP solution is approximated using the max-product algorithm for *Loopy Belief Propagation*. The algorithm passes messages between adjacent nodes in the network defining the support for the label assignment at each node. At each iteration, a node i passes a message m_{ij} to each neighbour $j \in N_i$ defining the belief in the label l_j given the label l_i. Messages are passed in parallel and the message $m'ij$ at the next iteration is defined by the update rule.

$$m'_{ij}(l_j) = \max_{l_i} \Psi(l_i, l_j) \Phi(l_i) \prod_{k \in N_i} m_{ki} \quad (12)$$

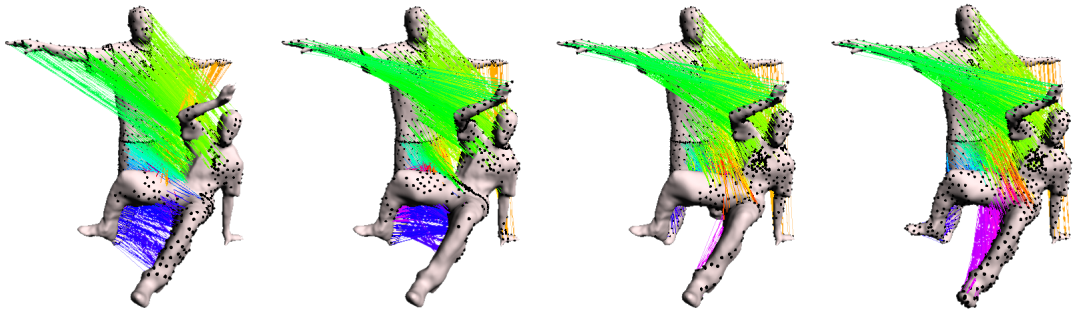
Belief propagation is performed for a set number of iterations and the label that maximises the final belief is selected at each node. The algorithm is initialised with no prior information on the consistency between nodes by setting $m^0_{ij} = 1$.

$$l_i = \arg\max_{l_i} \Phi(l_i) \prod_{j \in N_i} m_{ji} \quad (13)$$

Belief propagation in surface point matching is inherently ambiguous. Points on the different regions of the body such as the arms, torso, head and legs will have a high likelihood in matching all points in the same target region. Incorrect matches will also be supported by the MRF neighbourhood structure where neighbouring points have a high likelihood in incorrectly matching the neighbouring points in the target region. Similarly edge points at the boundary between regions can match any edge point on the target boundary. Only corner features provide a relatively distinct descriptor suitable for unambiguous matching but this occurs only at a sparse set of points as shown in Figure 3. The solution can therefore converge to a local optimum as shown in Figure 4(a).

A sparse-to-dense labelling strategy is introduced in which corner features are matched first to achieve surface registration, then edge and region feature points are subsequently matched. A process of cross-validation and pruning is introduced to ensure that only valid feature correspondences persist in belief propagation. Inference is performed in an MRF constructed from the set of features for matching, first {*corners*}, then {*corners,edges*}, and finally {*corners,edges,regions*}. Belief propagation in the MRF is performed using only a cross-validated set of labels to form the message updates m'_{ij}. The final set of labels are then extracted and pruned to a self-consistent set. Belief propagation in the MRF is then repeated using the consistent set of labels to initialise the messages m^0_{ij}.

Cross-validation is performed by verifying a consistent source-target, target-source labelling of correspondence. The posterior probability distribution is maintained at each



(a) Belief Propagation (b) Sparse to Dense (c) Validated (d) Regularised

Figure 4. Surface feature points showing extracted corners, edges and region points.

node in the network during belief propagation. This encodes an assignment matrix between the source and target feature points. Cross-validation is performed by only propagating messages where belief is maximised with a consistent source-to-target and target-to-source assignment.

Pruning is performed using a voting strategy to define the consistency in the final set of labels extracted from belief propagation. For each pair of feature points $d(x_i, x_j) < D_{max}$ consistency is voted where the geodesic surface distance on the source and target surface agree to within a threshold kD_{max}. A self-consistent set of labels is then extracted by pruning labels with a low number of votes.

The effect of belief propagation in deriving the set of correspondence labels L is illustrated in Figure 4. Belief propagation alone, Figure 4(a), results in a locally optimal solution due to the inherent ambiguity in matching the point descriptors. Sparse-to-dense matching improves correspondence for the arms but fails to correctly match the legs, Figure 4(b). Validation and consistency checking ensures that correspondence is correctly labelled across the whole body as shown in Figure 4(c). The proposed method serves to provide a validated set of consistently labelled surface correspondences.

4.3. Regularised likelihood

Complete surface matching is finally introduced using the prior shape of the model to regularise the labelling likelihood in the MRF. This is based on a common framework for shape matching in which an iterative process of correspondence estimation and model transformation is used to consistently label correspondence.

Regularisation is performed by iteratively deriving the labels L in belief propagation and deforming the surface to match the estimated correspondence, $l_i : x_s = y_t$. The surface is treated as an elastically deformable model that is optimised to minimise the distance between the surfaces $\underline{x} \in S$ and $\underline{y} \in T$ while preserving surface distance. The surface S is first updated to satisfy the rigid body transformation defined by the labels L and then optimised to satisfy the following deformation equation where x^0 denotes the original surface configuration for S.

$$S = \min_{\underline{x}} \sum_i \alpha \|\underline{y}_{l_i} - \underline{x}_i\|^2 + (1-\alpha)\frac{1}{N_i}\sum_{j \in N_i} (d(x_i, x_j) - d(x_i^0, x_j^0))^2 \quad (14)$$

In practise, non-linear geodesic surface distance is replaced by a linear Euclidean distance and the surface is optimised by gradient descent. The transformed feature point locations $\underline{x}_s$ are then used to derive the likelihood potential.

$$\Phi(l_i) = \exp\{-\lambda_F d(F_s, F_t)\} \exp\left\{-\lambda_P \|\underline{y}_t - \underline{x}_s\|^2\right\} \quad (15)$$

The influence of position regularisation is controlled by a parameter λ_P defining the expected variance from the predicted surface location. The effect of a regularised likelihood in the MRF is shown in Figure 4(d). A complete surface correspondence is now derived.

5. Evaluation

In this section, a quantitative evaluation of matching performance is presented for a set of surfaces from a public database of human surface motion [22]. A total of 25 frames were selected from 8 motions to test correspondence for a wide variation in body pose and across different motion sequences with changes in surface topology. The frames used in this evaluation are shown in Figure 5. A set of 25 feature points were manually marked on each surface on the head, arms, legs and torso, to provide a ground-truth estimate of correspondence.

A *true-positive* correspondence is defined where the labelled match for a marked feature point falls within a given *accuracy* threshold on geodesic surface distance. The true-positive rate, or *recall*, in matching is defined as the proportion of correctly labelled correspondence to the total



Figure 5. 25 Frames of 3D surface capture used to evaluate correspondence labelling. The topology is genus-0 (10 frames), 1 (8 frames), 2 (3 frames), 3 (2 frames) and 5 (2 frames).

number of correspondences. In matching all frames to all frames, the potential number of correspondences totals 15625 and in matching from one to all 625. Note that the *false-positive* rate, or *one-precision*, is not considered as the set of marked points is sparse leading to an artificially high precision. Instead the *recall - accuracy* characteristic is presented to evaluate the accuracy in localising matches.

Figure 6 shows the accuracy in matching from all frames to all frames. The recall is approximately 90% for an accuracy threshold of 20cm. It is important to note that the accuracy in manually labelling correspondence is non-zero, although this has not been quantified it is feasible that the error is in the order of 5 - 10cm. Performance is compared with point matching using the RPM-TPS algorithm [4], an alternating process of robust point matching (RPM) and surface deformation using a thin-plate spline (TPS). A comparison is also made against a graph matching technique [24], termed RL-RPM, in which the RPM-TPS algorithm is augmented with a relaxation labelling step that maintains the neighbourhood structure for the points.

Results demonstrate an improved performance for the proposed technique in Figures 6, 7 and Figure 8 shows a direct comparison of the recovered correspondence. The principal drawback in RPM-TPS and RL-RPM lies in the normalisation step for RPM which can break the neighbourhood consistency that is explicitly maintained in belief propagation with an MRF. The accuracy in matching a single reference frame to all frames is shown in Figure 7. The accuracy is higher for a genus-0 surface, with 90% correct labels within 10cm and almost 100% recall at 30cm. This highlights the fact that a genus-0 surface can more easily preserve local geodesic distance in matching. Matching from high to low genus will cause a large change in local surface distance. For example with the genus-1 surface shown in Figure 7(right), the hand touches the body giving

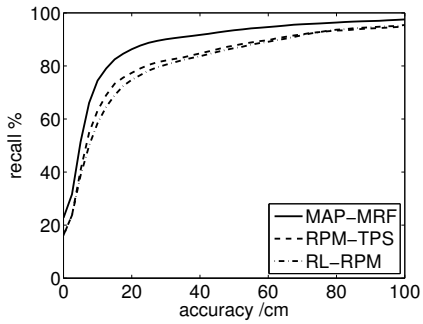


Figure 6. Recall-accuracy in matching all frames to all frames.

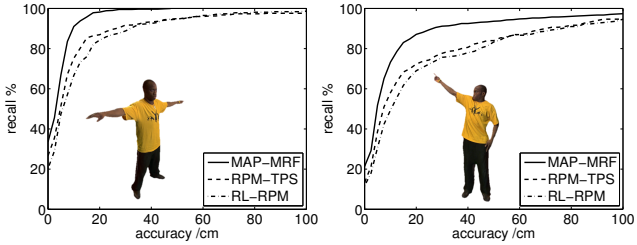


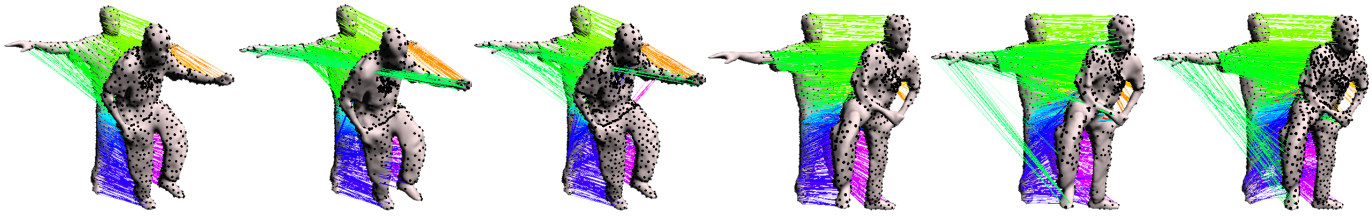
Figure 7. Recall-accuracy in matching (left) a genus-0 reference pose and (right) a genus-1 pose to all frames.

a local distance to be preserved in labelling that does not exist on the genus-0 surface shown in Figure 7(left).

6. Conclusions

In this paper the problem of estimating dense surface correspondence in arbitrary frames of captured surfaces has been addressed. The work focuses on matching the shape of a moving person with loose clothing where the match for corresponding parts of the body is highly ambiguous. A set of surface point descriptors have been introduced that are invariant to isometric surface deformations to construct a locally isometric surface mapping. A global surface match is achieved using a novel MAP-MRF framework that preserves the relative geodesic surface distance between points in labelling the most likely point correspondence. Matching is performed in a sparse-to-dense algorithm that leverages distinct corner features for surface registration and then progressively matches ambiguous edge and region features. Cross-validation and self-consistency is introduced to ensure that belief propagation does not converge to a local optimum and likelihood regularisation is introduced to derive a complete surface correspondence.

The performance of surface matching has been evaluated using surface sequences from a database of human motion. Ground-truth feature correspondences were manually marked and the recall-accuracy characteristic evaluated. Results demonstrate an improved performance compared to non-rigid point-pattern matching techniques, with



(a) MAP-MRF (b) RPM-TPS (c) RL-RPM (d) MAP-MRF (e) RPM-TPS (f) RL-RPM

Figure 8. Comparison of surface correspondence for the proposed MAP-MRF technique with RPM-TPS [4] and RL-RPM [24].

successful matching across wide variations in human body pose and surface topology. Correspondence labelling provides a promising mechanism to achieve wide-timeframe surface matches without the requirement for robust long-term sequential tracking or a prior model that restricts the space of feasible surface deformation. Further work is required to evaluate the attributes and descriptors adopted for surface correspondence and to validate the framework on different data-sets.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24):509–522, 2002. 2

[2] R. Carceroni and K. Kutulakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3d motion, shape and reflectance. *International Journal of Computer Vision*, 49(2-3):175–214, 2002. 2

[3] J. Carranza, C. M. Theobalt, M. Magnor, and H. Seidel. Free-viewpoint video of human actors. *ACM Transactions on Graphics (ACM SIGGRAPH 2003)*, 22(3):569–577, 2003. 2

[4] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003. 2, 7, 8

[5] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1285–1295, 2003. 2, 3

[6] R. Gal, A. Shamir, and D. Cohen-Or. Pose oblivious shape signature. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):261–271, 2007. 2

[7] T. Gatzke, C. Grimm, M. Garland, and S. Zelinka. Curvature maps for local shape comparison. *Shape Modeling and Applications*, pages 244–253, 2005. 2

[8] N. Gelfand, N. Mitra, L. Guibas, and H. Pottmann. Robust global registration. *Eurographics Symposium on Geometry Processing*, pages 197–206, 2005. 3

[9] N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani. Three dimensional shape searching: State-of-the-art review and future trends. *Computer-Aided Design*, 37(5):509–530, 2005. 2

[10] B. Levy, S. Petitjean, N. Ray, and J. Maillot. Least squares conformal maps for automatic texture atlas generation. *ACM Transactions on Graphics (ACM SIGGRAPH 2002)*, pages 362–371, 2002. 3

[11] H. Ling and D. Jacobs. Deformation invariant image matching. *IEEE International Conference on Computer Vision (ICCV)*, pages 1466–1473, 2005. 2, 4

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 4

[13] T. McInerney and D. Terzopoulos. Deformable models in medical images analysis: a survey. *Medical Imaging Analysis*, 1(2):91–108, 1996. 2

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005. 1, 2

[15] T. Moeslund, A. Hilton, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006. 2

[16] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, 2007. 2

[17] D. Pritchard and W. Heidrich. Cloth motion capture. *Computer Graphics Forum*, 22(3):263–271, 2003. 2

[18] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for non-rigid 3–d shape recovery. *to appear IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007. 2

[19] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor. Garment motion capture using color-coded patterns. *Computer Graphics Forum*, 24(3):439–448, 2005. 2

[20] J. Starck and A. Hilton. Model-based multiple view reconstruction of people. *IEEE International Conference on Computer Vision (ICCV)*, pages 915–922, 2003. 1, 2

[21] J. Starck and A. Hilton. Spherical matching for temporal correspondence of non-rigid surfaces. *IEEE International Conference on Computer Vision (ICCV)*, pages 1387–1394, 2005. 2, 3

[22] J. Starck and A. Hilton. Surface capture for performance based animation. *IEEE Computer Graphics and Applications*, 27(3):21–31, 2007. 2, 6

[23] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):475–480, 2005. 1, 2

[24] Y. Zheng and D. Doermann. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):643–649, 2006. 2, 7, 8