Towards optimal non-rigid surface tracking - Supplementary material

Martin Klaudiny, Chris Budd, and Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey, UK
m.klaudiny, chris.budd, a.hilton@surrey.ac.uk

1 Frame-to-frame surface tracking

The proposed method for non-sequential traversal of input sequences using traversal tree T can be combined with any frame-to-frame tracking technique working with the defined input measurements $\{O_t\}_{t=1}^N$ (for notation refer to the main paper). Two different techniques are used for the evaluation of the concept.

1.1 Image-oriented surface tracking (IOST)

The approach presented in [1] is aimed at open surfaces captured by a narrow-baseline camera setup where the fields of view are significantly overlapping. The assumed magnitude of frame-to-frame surface motion is moderate but the achieved precision of alignment is high. This is suitable for facial performance capture or cloth capture.

Multi-view 3D matching of textured surface patches to $I_{t_j}^c$ provides initial motion estimates between frames t_i, t_j which are weakly constrained by the shape of G_{t_j}. Subsequent weighted Laplacian deformation regularises raw 3D vertex displacements and deforms the tracked M_{t_i} to M_{t_j}. The tracking error E_{IOST} for a particular frame-to-frame transition is represented by the average error of 3D matching across all surface patches.

The dissimilarity measure d_{IOST} for IOST is derived from a sparse set of strong features robustly tracked in $\{I_t^c\}_{t=1}^N$ by a linear predictor tracker [2]. The 3D trajectories of features are obtained by back-projection of 2D trajectories onto $\{G_t\}_{t=1}^N$. $d(t_i, t_j)$ represents Euclidean distance between 3D positions of the features which are rigidly aligned beforehand between respective frames. The known rigid alignment is used to initialise the full frame-to-frame tracking algorithm.

1.2 Geometry-oriented surface tracking (GOST)

The approach presented in [?] is aimed at closed surfaces captured by surrounding wide-baseline camera setup. The spacious capture volume allows large free-form motion of the surface between frames. The focus of approach is robustness for larger frame-to-frame non-rigid deformation (e.g. fast motion) rather than high tracking accuracy. The primary application is whole-body performance capture.

ICP fitting of rigid surface patches from G_{t_i} to G_{t_j} provides initial 3D displacements between frames t_i, t_j. These are combined with a sparse set constraints from matching SIFT features between images $I^c_{t_i}$ and $I^c_{t_j}$. Initial deformation of the mesh M_{t_i} using the combined set of constraints is again performed by a Laplacian scheme. M_{t_i} is deformed further to M_{t_j} in coarse-to-fine fashion based on the growing number of displacements coming from ICP fitting of gradually smaller patches. The tracking error E_{GOST} for a particular frame-to-frame transition is represented by the average length of 3D trajectories travelled by rigid patches during iterative ICP fitting.

The dissimilarity measure d_{GOST} for GOST differs from IOST because it is difficult to obtain stable 3D trajectories of any features due to the complexity and variety of surface motion in this scenario. The shape histogram is employed instead because the motion is generally associated with large shape change for whole-body data [3]. The volumetric histogram based on spherical partitioning of the 3D space is calculated for both G_{t_i}, G_{t_j}. $d(t_i, t_j)$ is a sum of squared differences between the histograms which are optimised to increase correlation between them. A side product of the optimisation is a rigid alignment between the unregistered meshes which discards overall pose of the surface for frame-to-frame non-rigid alignment.

2 Trade-off between drift and jumps

To analyse the trade-off between drift and jumps across the spectrum of trees, two measures representing each of them are evaluated over tree structure. The measure SPL (shortest path length) reflects the amount of potential drift in individual frames. The amount of drift in frame t_k is related to the dissimilarity accumulated along the path $n_r \to n_k$ in the traversal tree T_β where n_r is the root node. SPL is the sum of path lengths from n_r to all other nodes (Equation 1).

$$SPL = \sum_{\forall n_k \in T_\beta} \sum_{\forall (n_i, n_j) \in n_r \to n_k} D(i,j) \tag{1}$$

The measure CUT reflects the magnitude of potential alignment inconsistencies at the cuts given by the structure of T_β. A difference in drift accumulation between adjacent frames t_k, t_l is related to the dissimilarity accumulated along their individual paths $n_r \to n_k$, $n_r \to n_l$ in the traversal tree T_β. The extent of different error accumulation is described by the length of non-overlapping parts of both paths $(n_b \to n_k) \subset (n_r \to n_k), (n_b \to n_l) \subset (n_r \to n_l)$ where n_b is a branching node which the paths separate at. This is evaluated for all pairs of adjacent frames which are not linked directly by an edge in $\mathcal{E}$ of T_β: $\bar{\mathcal{E}} = \{(n_k, n_l) : (n_k, n_l) \notin \mathcal{E}, |t_k - t_l| = 1\}$. Equation 2 for CUT defines the total sum of non-overlapping sub-paths for all cuts created by T_β.

$$CUT = \sum_{\forall(n_k,n_l)\in\bar{\mathcal{E}}} \left(\sum_{\forall(n_i,n_j)\in n_b\rightarrow n_k} D(i,j) + \sum_{\forall(n_i,n_j)\in n_b\rightarrow n_l} D(i,j) \right) \tag{2}$$

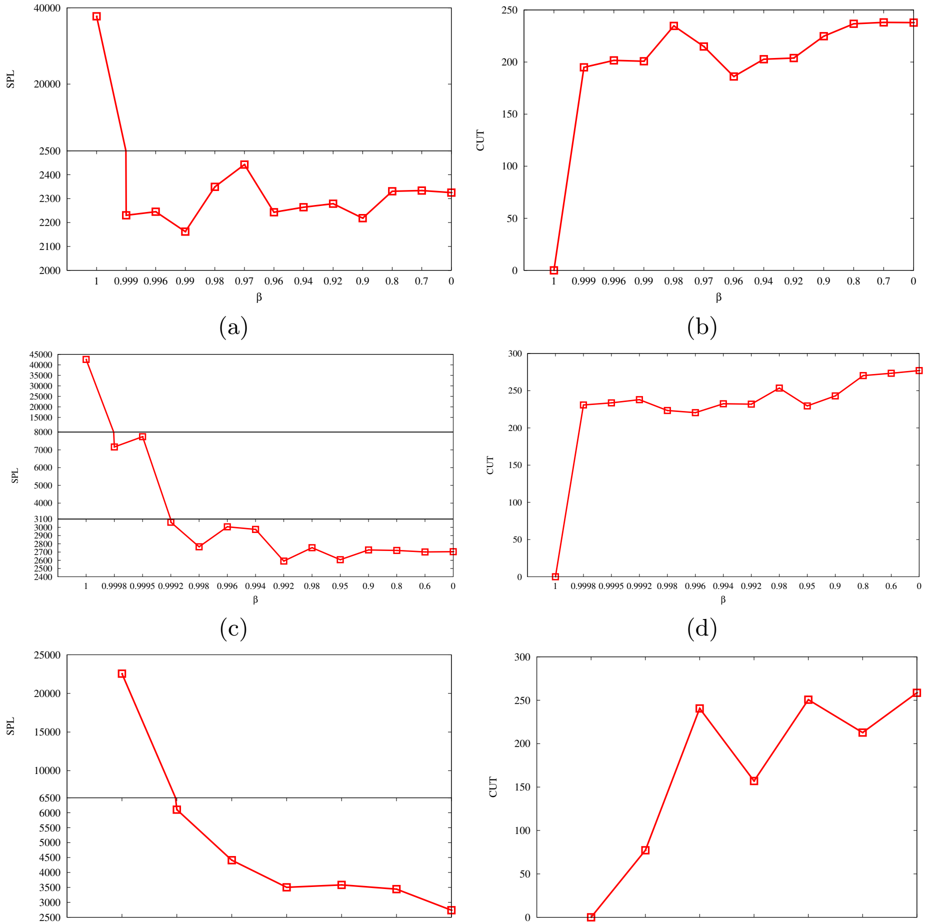
The profiles of SPL and CUT across different β are depicted in Figures 1, 2 for all datasets. Table 1 shows the number of clusters for the tested values of β across individual datasets (equivalent to the graph in Figure 3 in the main paper). Note that $\beta = 0.0$ is equivalent to MST (the number of clusters equals the number of frames in the sequence) and $\beta = 1.0$ is equivalent to purely sequential traversal (1 cluster).

Table 1. The values of β (with respective numbers of frame clusters) sampled for the cluster tree calculation across the datasets. β^* corresponds to the tree which gives the visually best tracking outcome.

Dataset	β(No. of clusters)	β^*
SyntheticFace	1.0(1), 0.999(23), 0.996(31), 0.99(41), 0.98(55), 0.97(61), 0.96(69), 0.94(83), 0.92(87), 0.9(97), 0.8(135), 0.7(175), 0.0(355)	0.99
Face	1.0(1), 0.9998(11), 0.9995(15), 0.9992(21), 0.998(25), 0.996(31), 0.994(35), 0.992(41), 0.98(59), 0.95(79), 0.9(99), 0.8(139), 0.6(217), 0.0(355)	0.95
DisneyFace	1.0(1), 0.999(15), 0.996(27), 0.99(37), 0.96(61), 0.9(91), 0.0(346)	0.996
Garment	1.0(1), 0.999(18), 0.997(30), 0.994(40), 0.98(60), 0.96(80), 0.92(104), 0.9(118), 0.8(188), 0.0(320)	0.994
StreetDance	1.0(1), 0.999(22), 0.998(30), 0.996(42), 0.994(52), 0.99(62), 0.97(102), 0.95(130), 0.93(152), 0.9(176), 0.8(250), 0.6(328), 0.0(1050)	0.996

References

1. Klaudiny, M., Hilton, A.: Cooperative patch-based 3D surface tracking. In: CVMP, Ieee (2011) 67–76
2. Ong, E.J., Lan, Y., Theobald, B.J., Harvey, R., Bowden, R.: Robust Facial Feature Tracking using Selected Multi-Resolution Linear Predictors. In: ICCV, IEEE (2009) 1483–1490
3. Huang, P., Hilton, A., Starck, J.: Shape Similarity for 3D Video Sequences of People. International Journal of Computer Vision **89** (2010) 362–381

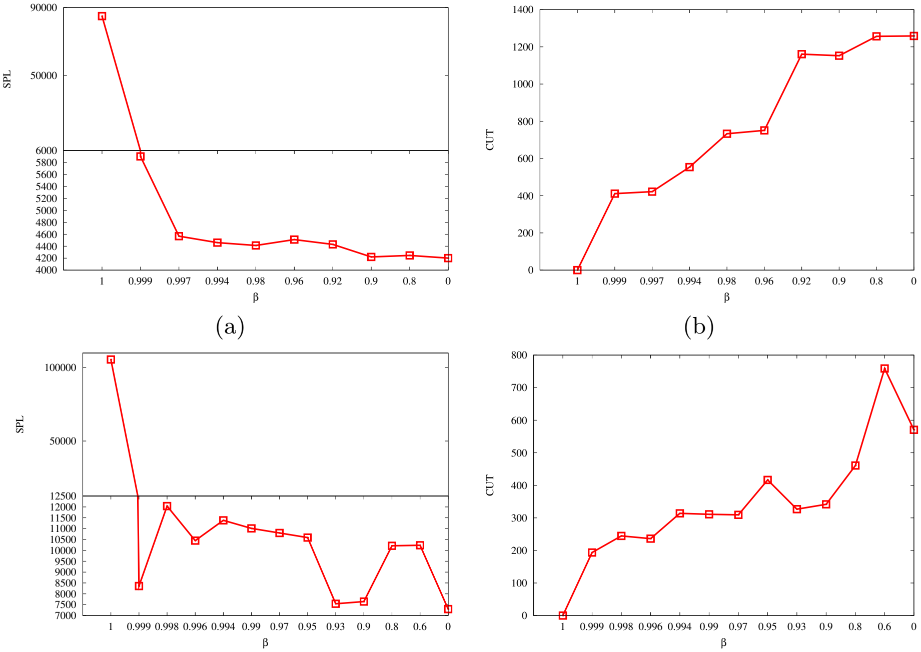


(a) (b)

(c) (d)

(e) (f)

Fig. 1. *SPL* and *CUT* measures across different traversal trees given by β for SyntheticFace (a,b), Face (c,d) and DisneyFace (e,f).



(a) (b)

(c) (d)

Fig. 2. SPL and CUT measures across different traversal trees given by β for Garment (a,b) and StreetDance (c,d).