VolumeDeform: Real-time Volumetric Non-rigid Reconstruction — Supplemental Document —

Matthias Innmann[1] Michael Zollhöfer[2]
Matthias Nießner[3] Christian Theobalt[2] Marc Stamminger[1]

[1]University of Erlangen-Nuremberg [2]Max-Planck-Institute for Informatics [3]Stanford University

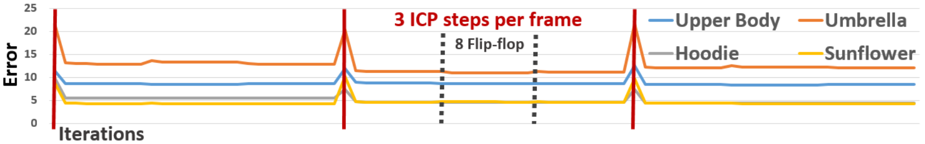

Fig. 1: Convergence of our approach.

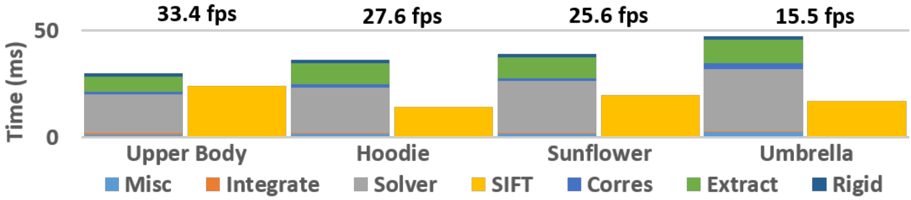

Fig. 2: Runtime performance of our approach.

In the following, we evaluate the runtime performance and convergence for four live sequences: UPPER BODY (1200 frames), HOODIE (552 frames), SUNFLOWER (876 frames), and UMBRELLA (727 frames). We use the thresholds $\epsilon_d = 3cm$ (distance), ϵ_n = 0.2 (normal deviation), and ϵ_v = 0.8 (view direction) for correspondence pruning and weighting. The influence of the different objectives is set to $w_s = 5$ (sparse), $w_d = 1$ (dense), and $w_r = 0.8$ (prior). Fig. 1 shows the convergence of the proposed optimization strategy on a single hierarchy level. Only for this evaluation, we perform 3 ICP steps with 8 flip-flop iterations (position and rotation). The linear system (position update) in each flip-flop iteration is solved based on 32 PCG steps. The largest improvement of residual error is achieved in the first ICP iteration step and in the first 4 flip-flop steps. Hence, as shown in the convergence graph, we only need 1-2 ICP steps and 4 flip-flop steps in practice (see below).

Fig. 2 shows the runtime performance of our approach for four live sequences. The figure visualizes the average per-frame runtime for all involved steps. Note that we distribute the work across two Nvidia GTX 980. The first card (GPU OPT) runs tracking and reconstruction (left bar), and the second card (GPU SIFT) performs SIFT feature extraction and matching (right bar). On average, GPU OPT runs for 38.2ms and GPU SIFT requires 18.7ms. We use 1-2 ICP steps (only UMBRELLA uses 2 steps), 4 flip-flop iterations, and 32 PCG steps. Our hierarchy has 3-4 levels (only UMBRELLA uses 3 levels). The number of grid points on the coarsest level is: UPPER BODY (2600 points), HOODIE (4000 points), SUNFLOWER (5500 points), and UMBRELLA (15k points). In this experiment, we compute the space deformation only on the coarsest level and upsample it to the resolution of the SDF.

List of Mathematical Symbols

Symbol	**Description**
$\mathcal{D}_t, \mathcal{C}_t$	depth, color input map at time t
$D, D^{-1}(0)$	distance field, zero level set (canonical pose)
$\mathcal{G}$	regular volumetric grid
$D_i, \mathbf{C}_i, W_i$	signed distance, color, weight at i-th grid point
$\hat{\mathbf{P}}$	polygonal mesh (canonical pose)
$\mathbf{P}$	polygonal mesh with space deformation applied
$\mathbf{R}_i, \mathbf{t}_i$	rotation, position of i-th grid point
$\hat{\mathbf{R}}_i, \hat{\mathbf{t}}_i$	rotation, position of i-th grid point (canonical pose)
$\mathbf{R}, \mathbf{t}$	global rotation, translation
$\mathbf{x}$	point in space
α_i	tri-linear interpolation weights
$\mathcal{S}$	space deformation
$\mathbf{p}_c, \mathbf{n}_c$	sample point, normal on $\mathbf{P}$
$\mathbf{p}_c^a, \mathbf{n}_c^a$	correspondence point, normal to $\mathbf{p}_c$ (input)
$\mathbf{v}$	current view direction
$\epsilon_d, \epsilon_n, \epsilon_v$	distance, normal, view threshold
$\phi_r(x)$	weighting kernel
w_c	confidence weight of c-th dense correspondence
$\mathbf{f}_s$	feature point
$\hat{\mathbf{f}}_s$	feature point in canonical pose
C, S	number of dense, sparse correspondences
$\mathcal{N}_i$	1-ring neighbourhood of i-th grid point
$\mathbf{X}, \mathbf{X}^*$	vector of all variables, optimal solution
$E_{total}, E_{sparse}, E_{dense}, E_{reg}$	total, sparse, dense, regularization objective
w_s, w_d, w_r	sparse, dense, regularization weight
$\mathcal{M}$	iso-surface plus 1-ring
N	number of grid points in $\mathcal{M}$
$\mathbf{L}$	Laplacian matrix
$\mathbf{B}$	constraint matrix
$\mathbf{b}$	right hand side of linear position system
$\mathbf{t}$	vector of all $\mathbf{t}_i$
K_{min}	time integration threshold