DispVoxNets: Non-Rigid Point Set Alignment with Supervised Learning Proxies

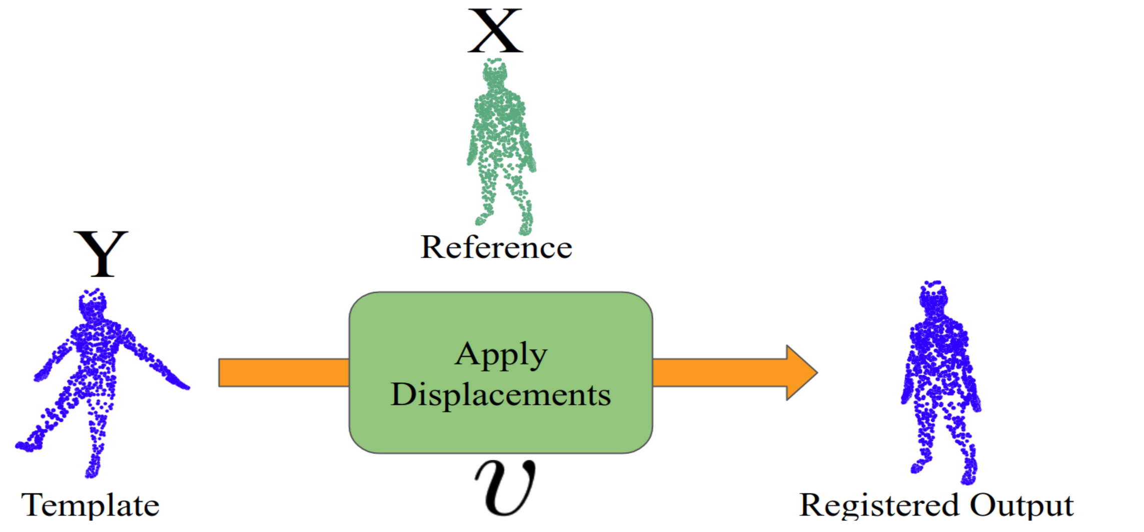
Soshi Shimada[1,2] Vladislav Golyanik[3] Edgar Tretschk[3] Didier Stricker[1,2] Christian Theobalt[3]

[1]University of Kaiserslautern [2]DFKI [3]Max Planck Institute for Informatics, SIC

3DV 2019 International Conference on 3D Vision Québec City, Canada September 16 - 19, 2019

Non-Rigid Point Set Registration

- Recovery of a displacement field aligning template (**Y**) and reference (**X**) point sets as well as correspondences between those.

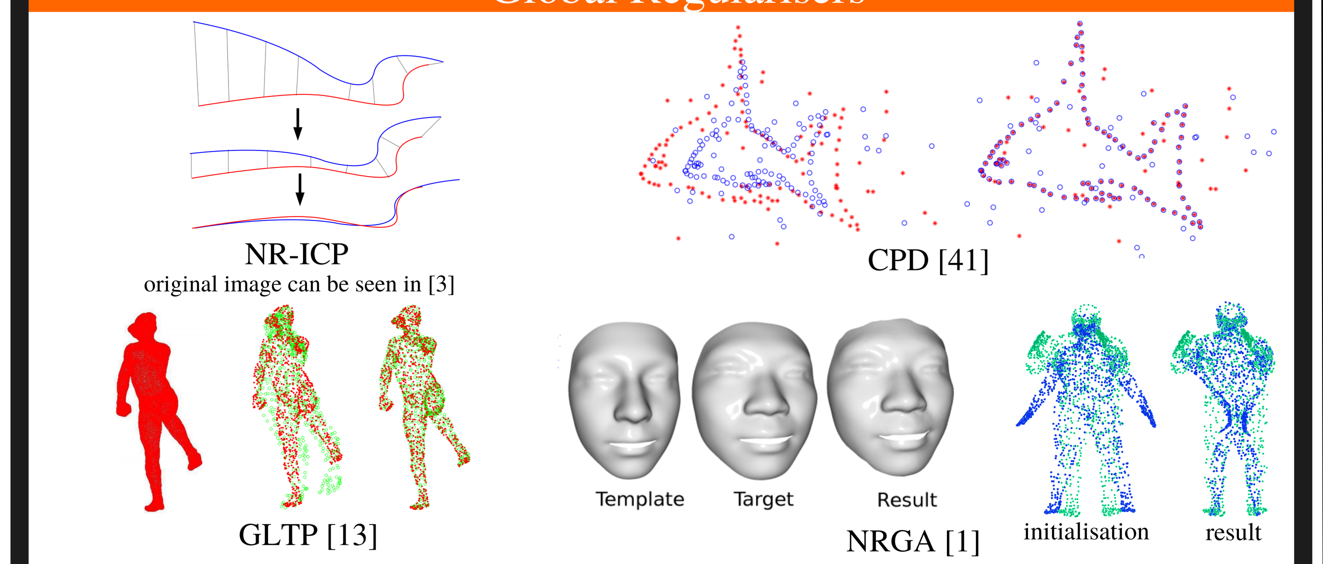


Motivation & Contributions

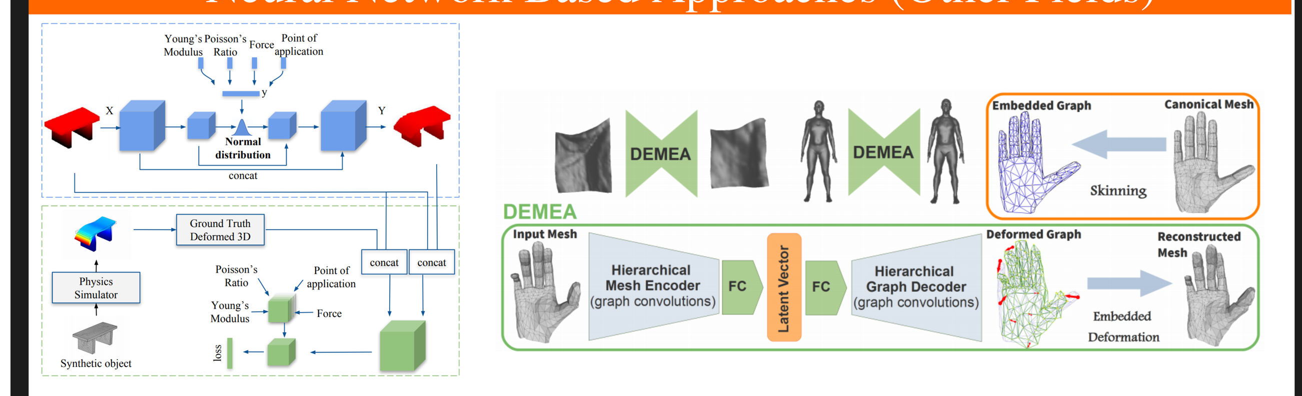
- We propose the **first neural network based general-purpose NRPSR method** that is invariant to the number and order of points.
- Our approach is **robust to large deformations, articulations, noises, outliers and missing data**.
- Our approach runs **orders of magnitude faster** than previous techniques.

Related Works

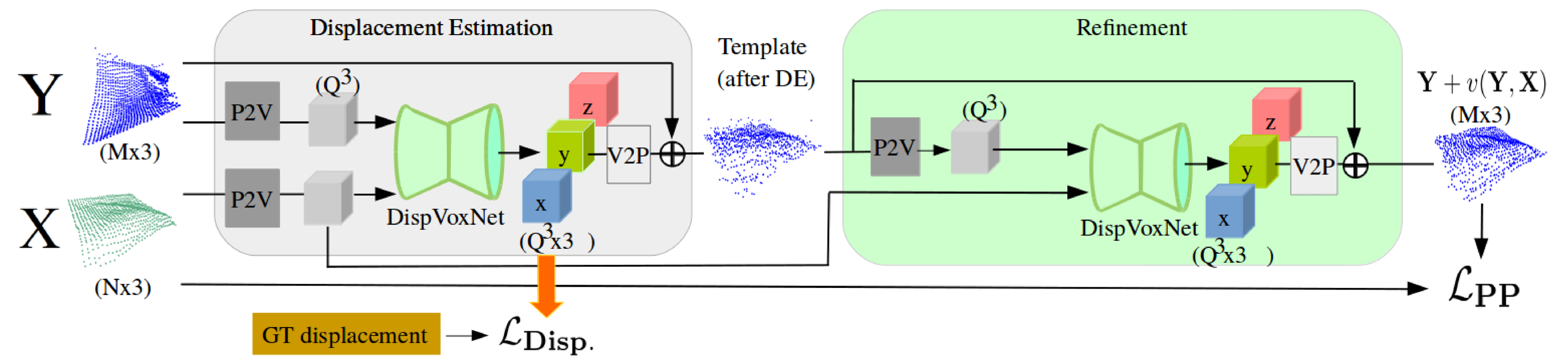
Global Regularisers



Neural Network Based Approaches (Other Fields)



DispVoxNets Pipeline



- Our pipeline is composed of two stages: Displacement Estimation (DE) and Refinement stage.
- DE stage regresses global displacements between Y and X.
- Refinement stage improves the initial displacements.

Y: template X: reference Q: size of the voxel grid P2V: point-to-voxel conversion
M: number of points in Y N: number of points in X V2P: voxel-to-point conversion
υ: displacement function

Loss Functions

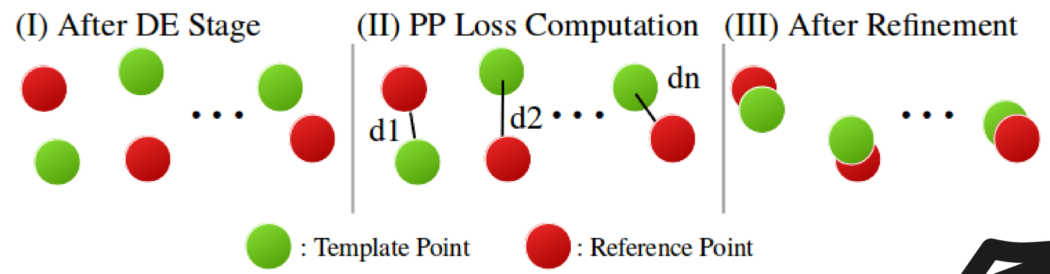
- ***Displacement loss*** penalises the discrepancy between the output displacements and ground truth displacements.

$$\mathcal{L}_{\text{Disp.}}(\mathbf{Z},\mathbf{V_Y},\mathbf{V_X}) = \frac{1}{Q^3}\|\mathbf{Z} - \mathbf{D_{vn}}(\mathbf{V_Y},\mathbf{V_X})\|_2^2$$

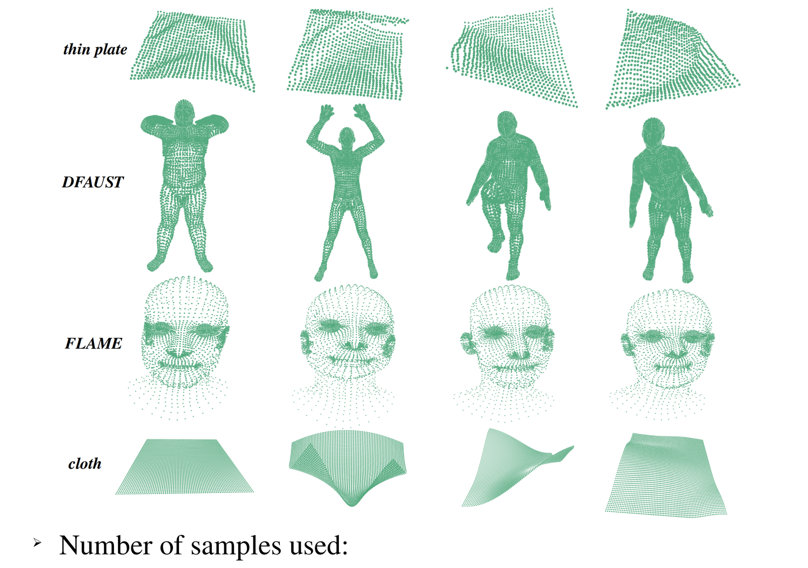
$\mathbf{V_Y}$: voxelised template $\mathbf{Z}$: GT displacements
$\mathbf{V_X}$: voxelised reference $\mathbf{D_{vn}}$: DispVoxNet

- ***Point projection loss*** penalises the Euclidean distances between a point y' in Y+υ(Y,X) and its closest point $\mathbf{x}_{\mathbf{y}'}$ in X.

$$\mathcal{L}_{\text{PP}}(\mathbf{Y} + \upsilon(\mathbf{Y},\mathbf{X}),\mathbf{X}) = \frac{1}{M}\sum_{i=1}^{M}\|\mathbf{y}'_i - \mathbf{x}_{\mathbf{y}'}\|_2$$

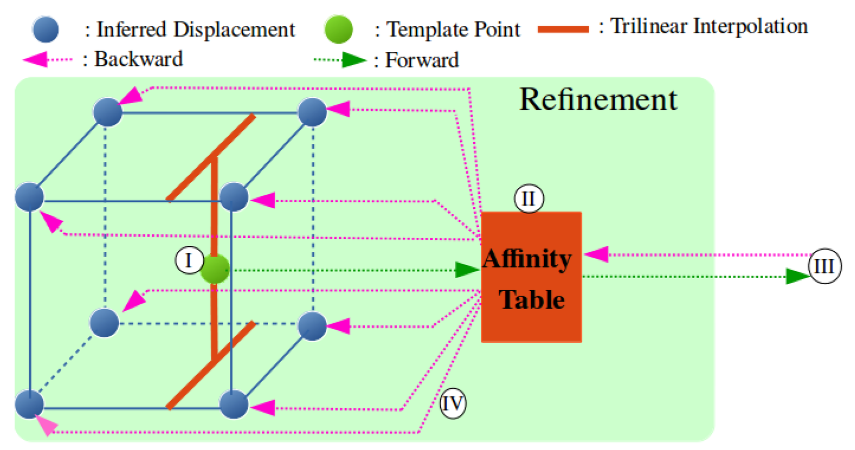


Datasets



- Number of samples used:
  thin plate [17]: 4,648 *FLAME* [36]: 10,000
  DFAUST [5]: 7,948 *Cloth* [2]: 6,238

Trilinear Weighting



I. Compute trilinear weights to estimate sub-voxel displacements.

II. Record the weights and indices of the 8 nearest displacements in the affinity table.

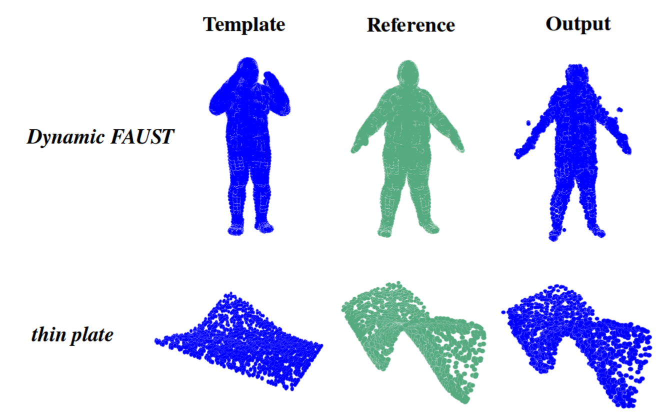
III. Compute the point projection loss.

IV. Distribute gradients following the IDs and weights information recorded in the affinity table in II.

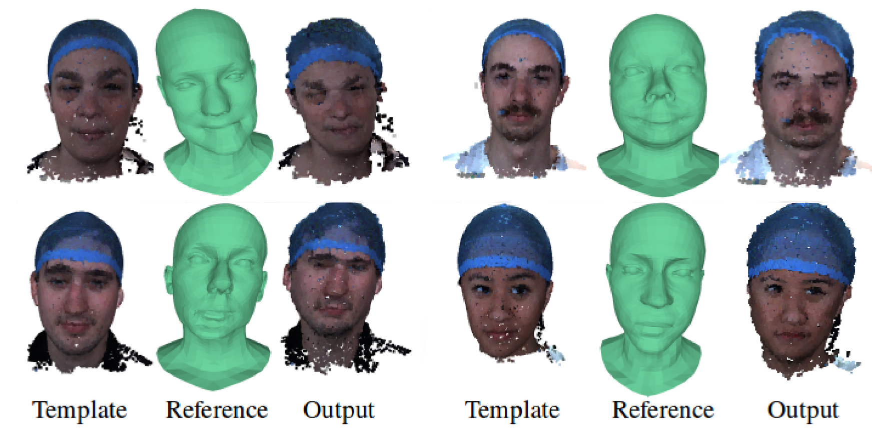
Results

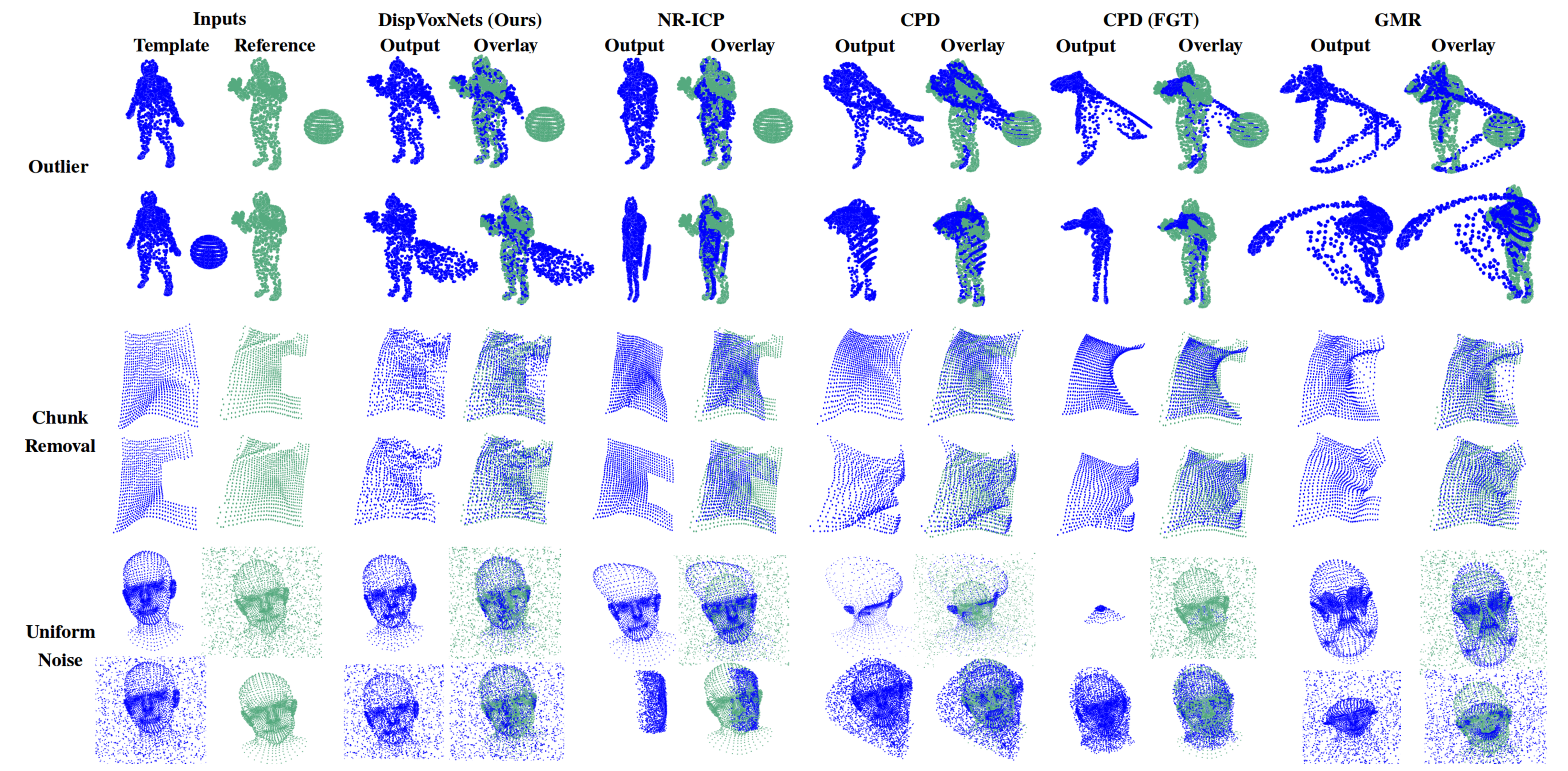
Qualitative Results

- Clean Data



- Real Scan Data





- Deteriorated Data

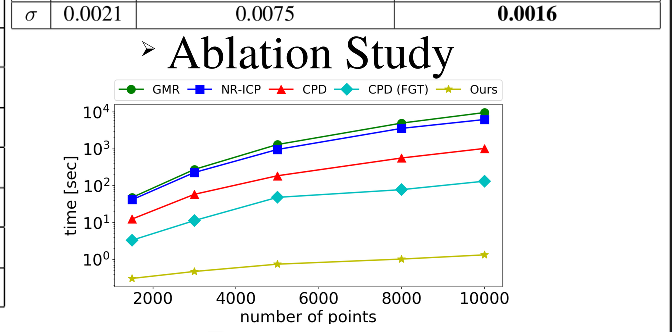
Quantitative Results

		Ours	NR-ICP [9]	CPD [41]	GMR [31]
thin plate[17]	e	0.0103	0.0402	**0.0083** / 0.0192	0.2189
	σ	**0.0059**	0.0273	0.0102 / 0.0083	1.0121
FLAME[36]	e	0.0063	0.0588	**0.0043** / 0.0094	0.0056
	σ	0.0009	0.0454	0.0008 / **0.0005**	0.0007
DFAUST[5]	e	**0.0166**	0.0585	0.0683 / 0.0721	0.2357
	σ	**0.0020**	0.0215	0.0314 / 0.0258	0.8944
cloth[2]	e	**0.0080**	0.0225	0.0149 / 0.0138	0.2189
	σ	**0.0021**	0.0075	0.0066 / 0.0033	1.0121

- Clean Data

	DE	DE + Ref. (nearest voxel)	Full: DE + Ref. (trilinear)
e	0.0100	0.0088	**0.0069**
σ	0.0021	0.0075	**0.0016**

- Ablation Study



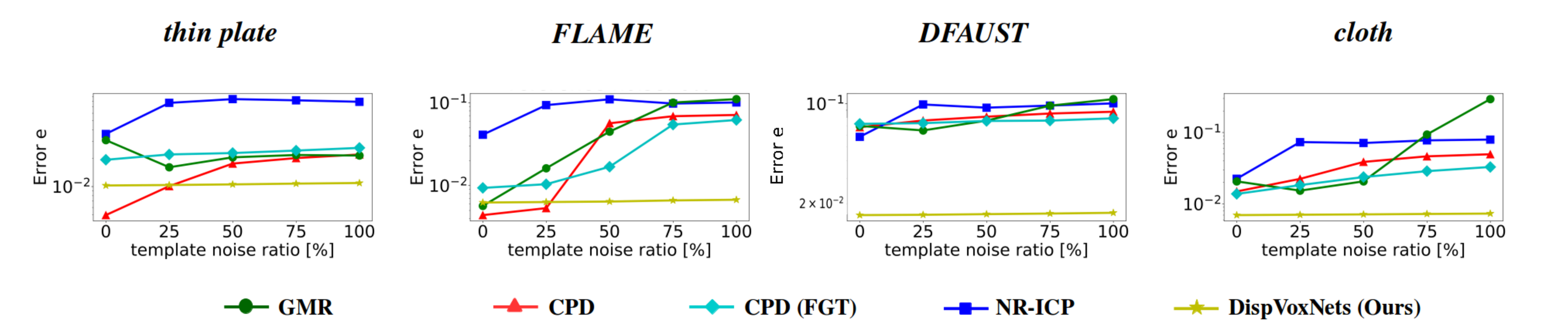
- Runtime

			Ours	NR-ICP [9]	CPD [41]	GMR [31]
thin plate[17]	ref.	e	**0.0107**	0.0668	0.0218 / 0.0386	0.4415
		σ	**0.0061**	0.0352	0.0148 / 0.0067	1.4632
	temp.	e	**0.0108**	0.0334	0.0479 / 0.0471	0.4287
		σ	0.0062	0.0281	0.0101 / **0.0038**	1.3832
FLAME[36]	ref.	e	0.0084	0.0519	**0.0046** / 0.0140	0.0193
		σ	0.0010	0.0451	0.0009 / **0.0006**	0.0008
	temp.	e	0.0088	0.0215	**0.0076** / 0.0201	0.0274
		σ	**0.0010**	0.0219	**0.0010** / 0.0016	0.0019
DFAUST[5]	ref.	e	**0.0167**	0.0463	0.0562 / 0.0636	0.0714
		σ	**0.0029**	0.0195	0.0308 / 0.0216	0.0282
	temp.	e	**0.0169**	0.0426	0.0672 / 0.0710	0.0737
		σ	**0.0033**	0.0194	0.0291 / 0.0229	0.0243
cloth[2]	ref.	e	**0.0090**	0.0455	0.0248 / 0.0315	0.0288
		σ	**0.0018**	0.0061	0.0056 / 0.0027	0.0087
	temp.	e	**0.0132**	0.0208	0.0486 / 0.0347	0.0397
		σ	0.0019	0.0087	0.0077 / **0.0014**	0.0092

- Missing Data

			Ours	NR-ICP [9]	CPD [41]	GMR [31]
thin plate[17]	ref.	e	**0.0151**	0.0349	0.1267 / 0.1136	0.6332
		σ	**0.0117**	0.0302	0.0224 / 0.0211	1.5749
	temp.	e	**0.0150**	0.0509	0.0304 / 0.0636	0.0528
		σ	**0.0106**	0.0406	0.0200 / 0.0149	0.0300
FLAME[36]	ref.	e	0.0098	**0.0039**	0.0492 / 0.0617	0.0577
		σ	0.0034	**0.0007**	0.0301 / 0.0218	0.0205
	temp.	e	0.0073	0.0566	**0.0072** / 0.0246	0.0309
		σ	**0.0015**	0.0334	0.0070 / 0.0142	0.0117
DFAUST[5]	ref.	e	**0.0308**	0.0605	0.1127 / 0.1151	0.9730
		σ	**0.0111**	0.0226	0.0308 / 0.0295	2.2267
	temp.	e	**0.0190**	0.0669	0.0791 / 0.0775	0.0845
		σ	**0.0036**	0.0187	0.0304 / 0.0220	0.0295
cloth[2]	ref.	e	**0.0213**	0.0248	0.1081 / 0.1096	0.1098
		σ	**0.0091**	0.0095	0.0235 / 0.0223	0.0234
	temp.	e	0.0649	**0.0296**	0.0408 / 0.0522	0.0476
		σ	0.0395	**0.0081**	0.0115 / 0.0114	0.0223

- Outlier



- Uniform Noise

References & QR Code

[1] S. A. Ali, V. Golyanik, and D. Stricker. Nrga: Gravitational approach for non-rigid point set registration. In *International Conference on 3D Vision (3DV)*, 2018.
[2] J. Bednarık, P. Fua, and M. Salzmann. Learning to reconstruct texture-less deformable surfaces from a single view.In *International Conference on 3D Vision (3DV)*, 2018.
[3] E. Smistad, et al. Medical image segmentation on GPUs– A comprehensive review. In *Medical Image Analysis*, 2015.
[5] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black. Dy-namic FAUST: Registering human bodies in motion. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.
[9] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. In *Computer Vision and Image Understanding (CVIU)*, 2003.
[13] S. Ge, G. Fan, and M. Ding. Non-rigid point set registration with global-local topology preservation. In *Computer Visionand Pattern Recognition Workshops (CVPRW)*, 2014.
[17] V. Golyanik, S. Shimada, K. Varanasi, and D. Stricker. HDM-Net: Monocular non-rigid 3d reconstruction with learned deformation model. In *EuroVR*, 2018.
[31] B. Jian and B. C. Vemuri. A robust algorithm for pointset registration using mixture of gaussians. In *International Conference for Computer Vision (ICCV)*, 2005.
[36] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero. Learn-ing a model of facial shape and expression from 4D scans. *SIGGRAPH Asia*, 2017.
[41] A. Myronenko and X. Song. Point-set registration: Coherentpoint drift. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010.
[58] E. Tretschk, A. Tewari, M. Zollhofer, V. Golyanik, and ¨ C. Theobalt. DEMEA: Deep Mesh Autoencoders for NonRigidly Deforming Objects. *arXiv e-prints*, 2019
[60] Z. Wang, S. Rosa, and A. Markham. Learning the intuitive physics of non-rigid object deformations. In *Neural Information Processing Systems (NIPS) Workshops*, 2018

Paper Supplementary arXiv Video Slides