Seminar Visual Computing
Final Presentation

NPMs: Neural Parametric Models for 3D Deformable Shapes

August 31, 2023

Bertan Karacora
Institute of Computer Science, University of Bonn
bertan.karacora@uni-bonn.de

Contents

Contents

Modeling 3D deformable shapes

Figure 1: Deformable shapes in comparison with rigidly transformable objects. Left: Beethoven statue at the Münsterplatz in Bonn [1]. Center: LEGO® figure of Beethoven [2]. Right: AI-generated image of Beethoven, created with StableDiffusion [3].

➢ How to account for these deformations?

Parametric models

- Control of distinct properties (e.g., template mesh, joint angles)
- PCA-based example: SMPL [5]

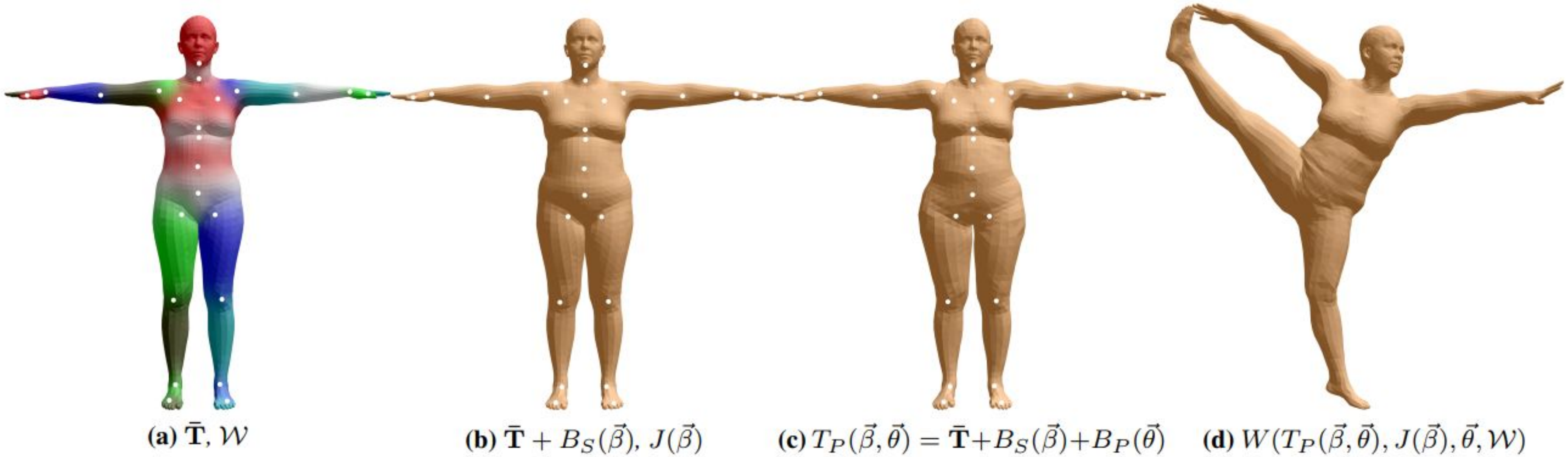


Figure 2: Overview of the SMPL model. Parameters like the template mesh, blend weights, and blend shapes are gained from statistical analysis. From [5].

➢ Parametrization is a hard constraint.

Motivating NPMs [6]

- Create parametric models for any domain without manual annotations
  - Disentangle shape and pose
  - Learn the parametrization from data
  - Leverage implicit representations of geometry and deformations

Contents

Overview

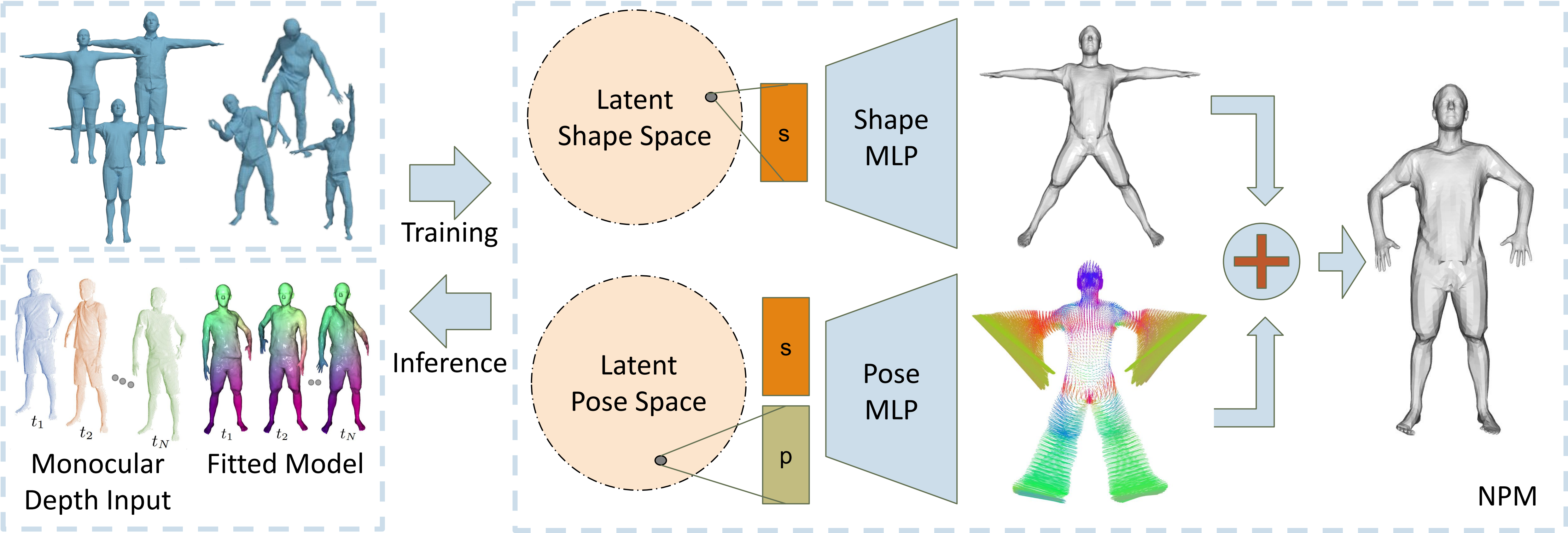


Figure 3: Overview of NPMs. NPMs learn latent spaces of shape and pose and optimize them jointly to fit new observations.

Auto-decoder training

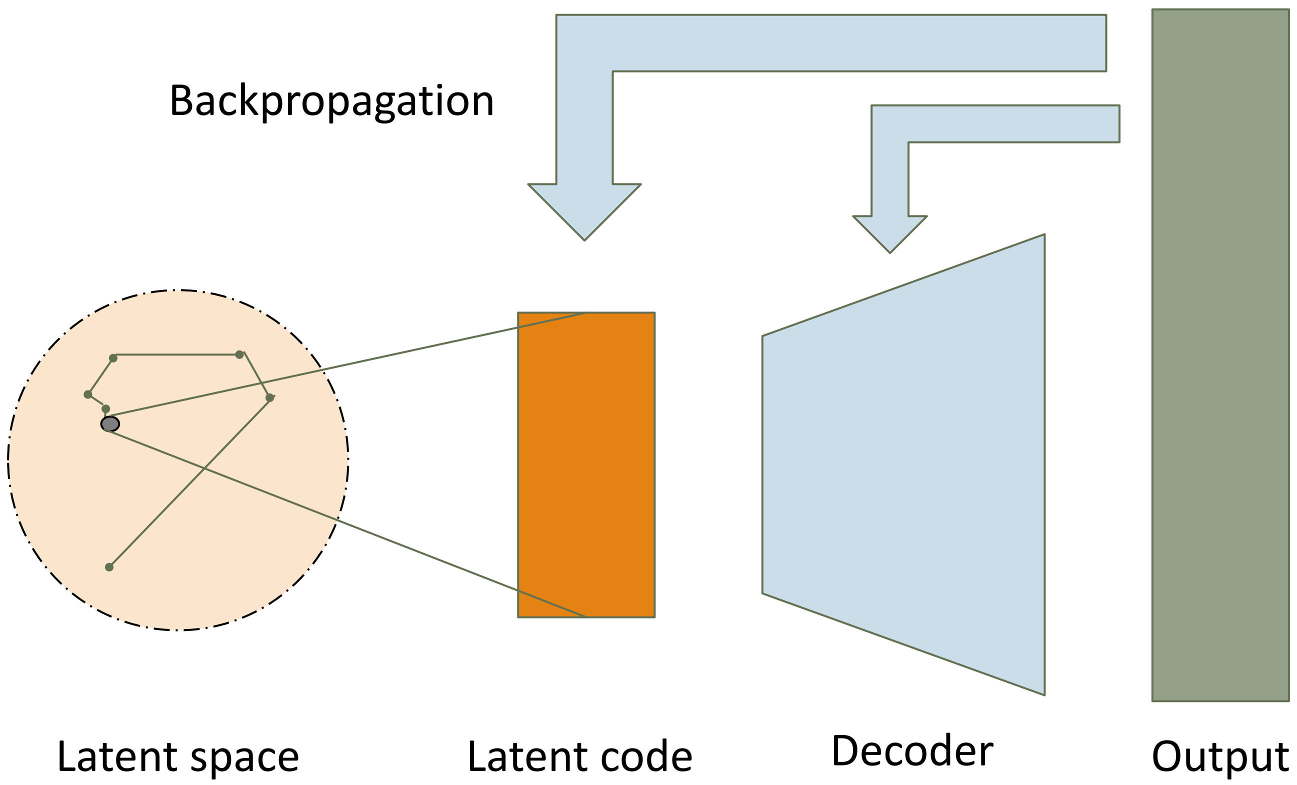


Figure 4: Auto-decoder learning.

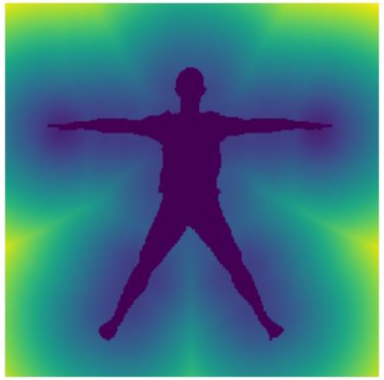
Latent shape space

- Space of canonically posed shapes
- Implicit representation as signed distance field:

$$f_{\theta_\mathrm{s}} : \mathbb{R}^{D_\mathrm{s}} \times \mathbb{R}^3 \to \mathbb{R},$$
$$(\boldsymbol{s}_i, \boldsymbol{x}) \mapsto f_{\theta_\mathrm{s}}(\boldsymbol{s}_i, \boldsymbol{x}) = \tilde{d}$$

- Reconstruction energy:

$$\underset{\theta_\mathrm{s}, \{\boldsymbol{s}_i\}_{i=1}^{S}}{\arg\min} \sum_{i=1}^{S} \left(\sum_{k=1}^{N_\mathrm{s}} \mathcal{L}_\mathrm{s}(f_{\theta_\mathrm{s}}(\boldsymbol{s}_i, \boldsymbol{x}_i^k), d_i^k) + \frac{\|\boldsymbol{s}_i\|_2^2}{\sigma_\mathrm{s}^2} \right)$$



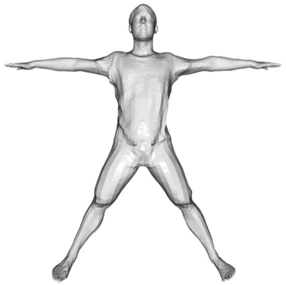


Figure 5: Truncated SDF slice (top) and constructed 3D mesh(bottom).

Latent pose space

- Space of valid poses of the shapes from shape space
- Implicit representation as surface deformations:

$$f_{\theta_\mathrm{p}} : \mathbb{R}^{D_\mathrm{s}} \times \mathbb{R}^{D_\mathrm{p}} \times \mathbb{R}^3 \to \mathbb{R}^3,$$
$$(\boldsymbol{s}_i, \boldsymbol{p}_j, \boldsymbol{x}) \mapsto f_{\theta_\mathrm{p}}(\boldsymbol{s}_i, \boldsymbol{p}_j, \boldsymbol{x}) = \Delta\tilde{\boldsymbol{x}}$$

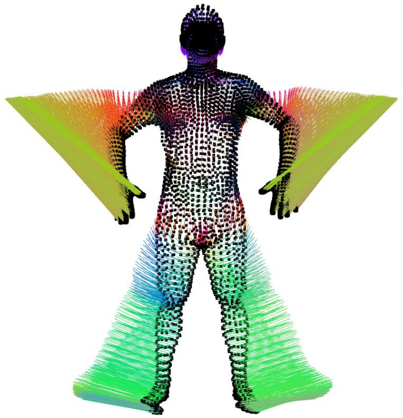


Figure 6: Deformations with regards to the posed body mesh vertices.

- Reconstruction energy:

$$\underset{\theta_\mathrm{p}, \{\boldsymbol{p}_j\}_{j=1}^{P}}{\arg\min} \sum_{\substack{j=1 \\ i=m[j]}}^{P} \left(\sum_{k=1}^{N_\mathrm{p}} \mathcal{L}_\mathrm{p}(f_{\theta_\mathrm{p}}(\boldsymbol{s}_i, \boldsymbol{p}_j, \boldsymbol{x}_i^k), \Delta\boldsymbol{x}_{ij}^k) + \frac{\|\boldsymbol{p}_j\|_2^2}{\sigma_\mathrm{p}^2} \right)$$

Overview

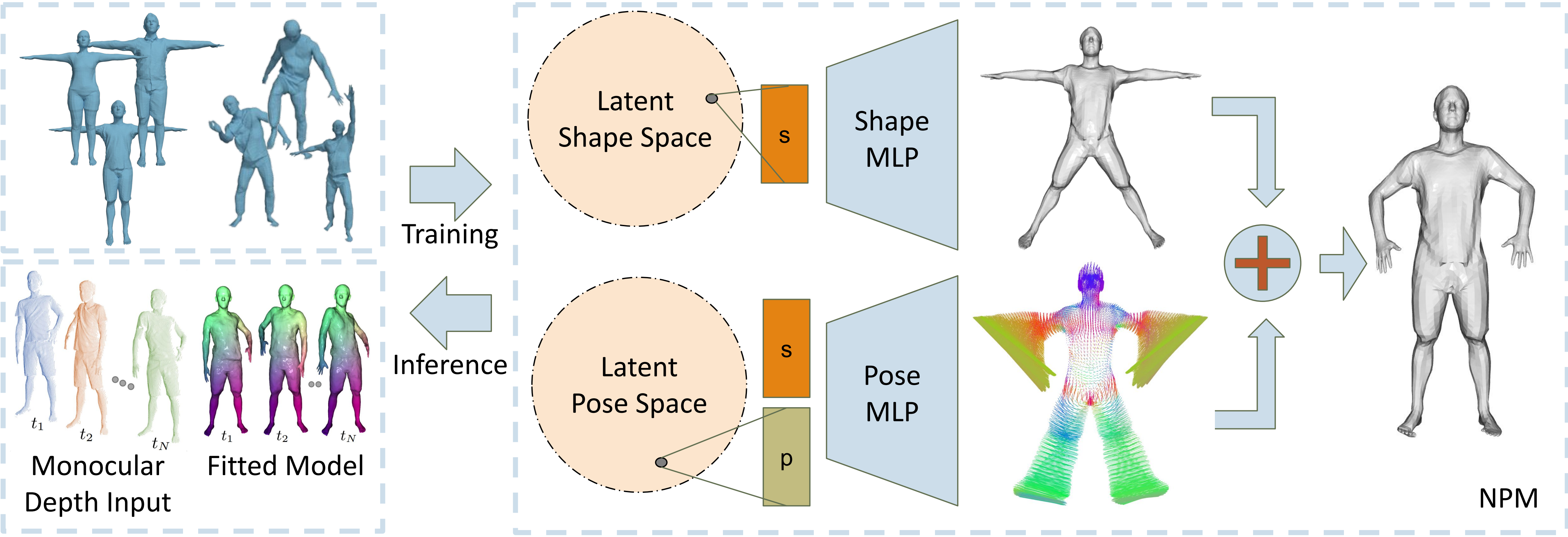


Figure 3: Overview of NPMs. NPMs learn latent spaces of shape and pose and optimize them jointly to fit new observations.

Test-time optimization for fitting

- Energy function:

$$\underset{\boldsymbol{s},\{\boldsymbol{p}_j\}_{j=1}^{L}}{\arg\min} \sum_{j=1}^{L} \sum_{\boldsymbol{x}_k} \mathcal{L}_{\mathrm{r}} + \mathcal{L}_{\mathrm{c}} + \mathcal{L}_{\mathrm{t}} + \mathcal{L}_{\mathrm{icp}}$$

- Reconstruction loss:

$$\mathcal{L}_{\mathrm{r}} = M_{\mathrm{o}} \mathcal{L}_{\mathrm{s}} \left(f_{\theta_s}(\boldsymbol{s}, \boldsymbol{x}_k), \left[\boldsymbol{x}_k + f_{\theta_{\mathrm{p}}}(\boldsymbol{s}, \boldsymbol{p}_j, \boldsymbol{x}_k) \right]_{\mathrm{sdf}} \right)$$

- Regularization
  - Gaussian priors
  - Temporal consistency
  - ICP-inspired alignment consistency

Contents

Fitting to human bodies

- Evaluation on CAPE dataset of clothed humans [7]
- Depth map sequence shows a fluent motion
- Similar observations for fitting to hands

Qualitative results

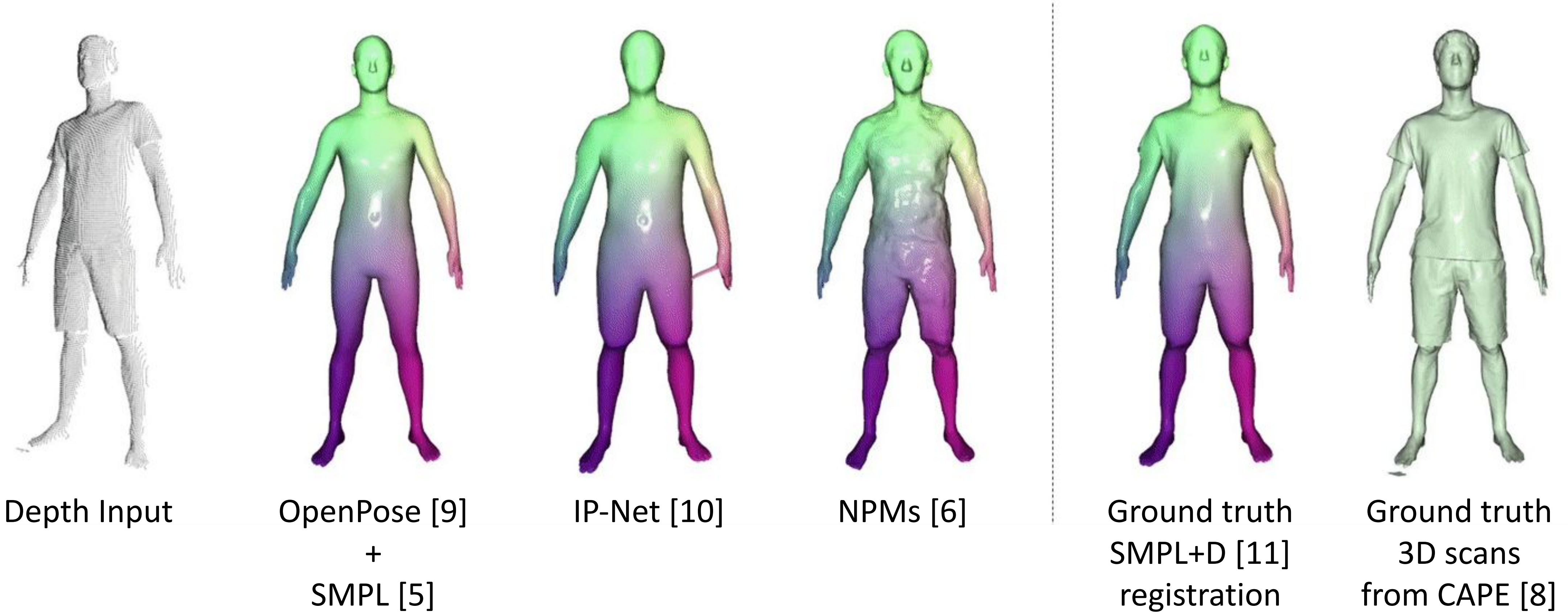


Figure 7: Qualitative results and comparison with state-of-the-art (at the time) models for non-rigid 4D reconstruction from monocular depth. Adapted from [6].

Latent space interpolation

Shape

Pose

Figure 8: Shape and pose interpolation with NPMs. From [6].

Shape and pose transfer

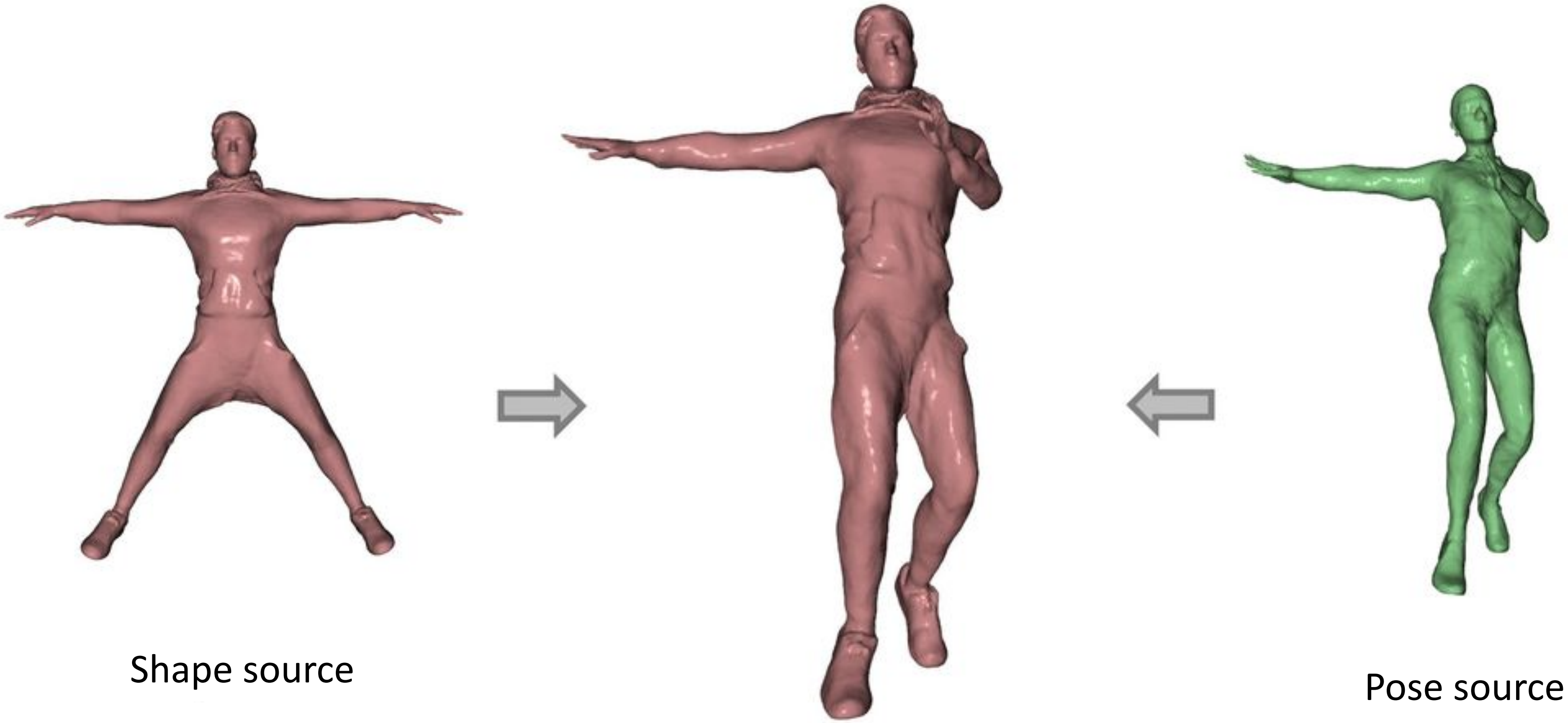


Figure 9: Shape and pose transfer with NPMs. From [6].

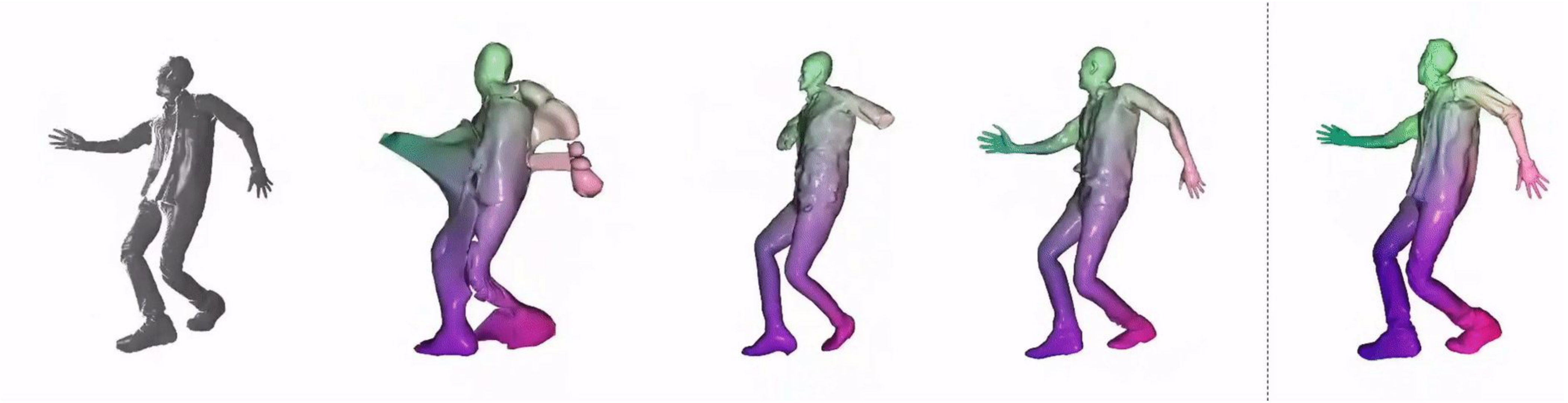
Contents

Limitations

- No semantic meaning of parameters
  - Lack of intuition and interpretability
  - Only indirect control/manipulation
  - Not suitable for generative tasks
- Does not account for perception (e.g., importance of faces)
- Limited expressiveness (e.g., loose clothing)
- Assumptions about training data
- Heavy computation

Further developments

- Structured NPMs (e.g., SPAMs [12], NPHMs [13])



Depth Input | IP-Net [10] | NPMs [6] | SPAMs [12] | GT SMPL+D [11] registration

Figure 10: Comparison with SPAMs. Adapted from [12].

➢ Re-adding handcrafted constraints (e.g., body structure, symmetry).

Contents

Conclusion

- Flexible approach for parametric model construction
  - Disentangling shape and pose
  - Learning implicit representations in auto-decoder fashion
- Accurate fitting and interpolation in learned spaces
- Lack of semantic meaning
  - Recent improvements over NPMs by re-adding handcrafted segmentation

References

In order of occurrence:

- [1] Hähnel, E (2008). Photo of the Ludwig van Beethoven statue at the Münsterplatz in Bonn, Germany. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Beethoven_monument_bonn_muensterplatz_2008.jpg. Accessed August 28, 2023.
- [2] Hobby Brick (2018). Photo of a Lego figure of Ludwig van Beethoven. https://www.herobloks.com/figures/17006/pinterest. Accessed August 28, 2023.
- [3] Rombach, R., A. Blattmann, D. Lorenz, P. Esser and B. Ommer (2021). “High-Resolution Image Synthesis with Latent Diffusion Models”. In: arXiv preprint arXiv:2112.10752.
- [4] Gafni, G., J. Thies, M. Zollhöfer, and M. Nießner (2021). “Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction”. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8649–8658.
- [5] Loper, M., N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black (2015). “SMPL: A Skinned Multi-Person Linear Model”. In: ACM Trans. Graphics (SIGGRAPH Asia) 34.6, 248:1–248:16.
- [6] Palafox, P., A. Božič, J. Thies, M. Nießner, and A. Dai (2021). “NPMs: Neural Parametric Models for 3D Deformable Shapes”. In: IEEE/CVF International Conference on Computer Vision (ICCV).
- [7] Park, J. J., P. Florence, J. Straub, R. Newcombe, and S. Lovegrove (2019). “DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [8] Anguelov, D., P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis (2005). “SCAPE: Shape Completion and Animation of People”. In: ACM SIGGRAPH 2005 Papers. SIGGRAPH ’05, pp. 408–416.
- [9] Cao, Z., G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh (2019). “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [10] Bhatnagar, B. L., C. Sminchisescu, C. Theobalt, and G. Pons-Moll (2020). “Combining Implicit Function Learning and Parametric Models for 3D Human Reconstruction”. In: European Conference on Computer Vision (ECCV).
- [11] Alldieck, T., M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. Pons-Moll (2019). “Learning to Reconstruct People in Clothing from a Single RGB Camera”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [12] Palafox, P., N. Sarafianos, T. Tung, and A. Dai (2022). “SPAMs: Structured Implicit Parametric Models”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [13] Giebenhain, S., T. Kirschstein, M. Georgopoulos, M. Rünz, L. Agapito, and M. Nießner (2023). “Learning Neural Parametric Head Models”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Additional slides

Applications

- 3D shape reconstruction and pose tracking
- Avatar creation
- Novel view synthesis
- Artificial re-animation
- …

Figure 11: Transfering facial expressions from one identity to another. Adapted from [4].

Pose representation

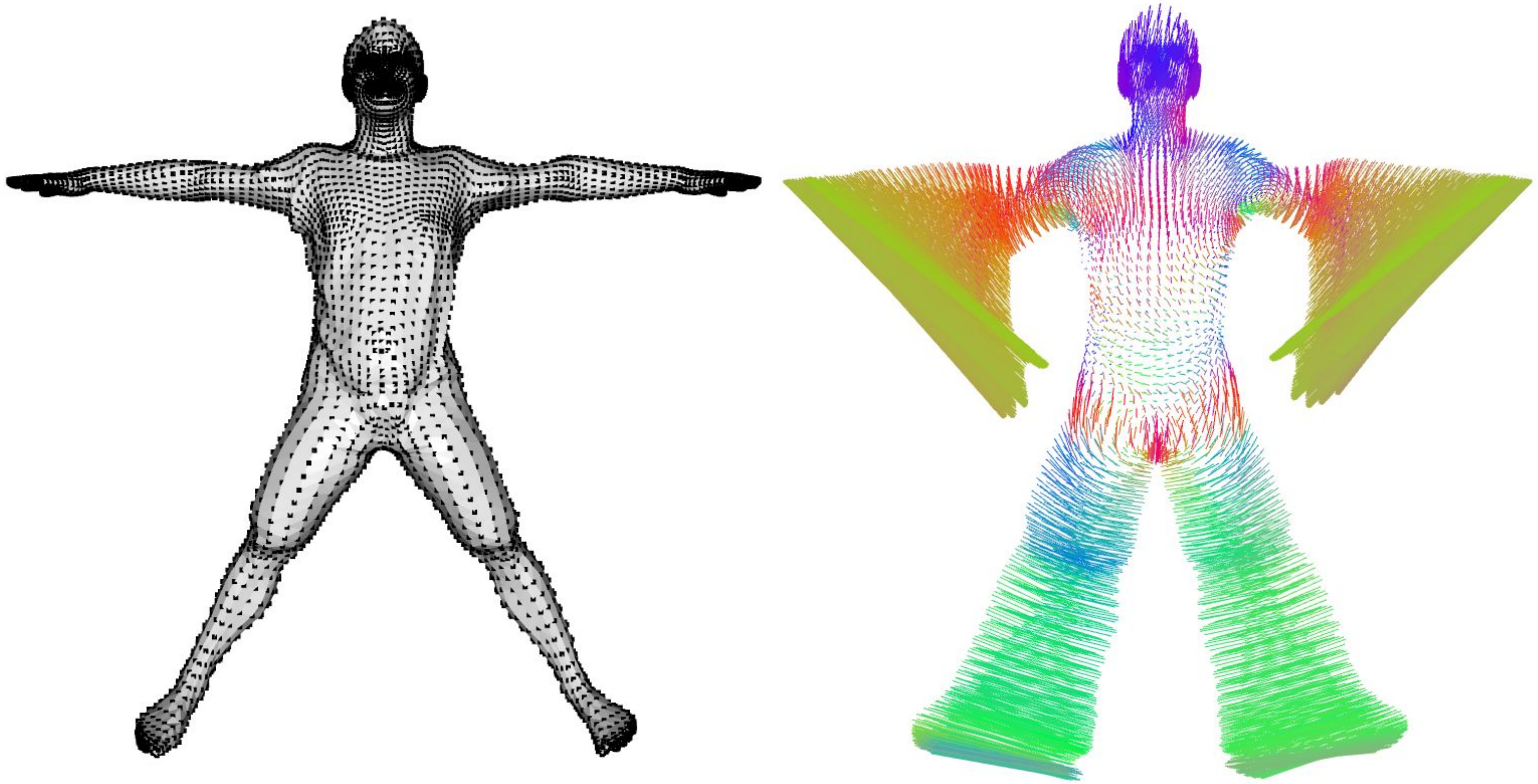


Figure 12: Implicit pose representation as flow vectors. Reconstruction is performed by sampling the pose function on the surface vertices and adding to the canonically posed mesh.

Monocular depth as partial SDF

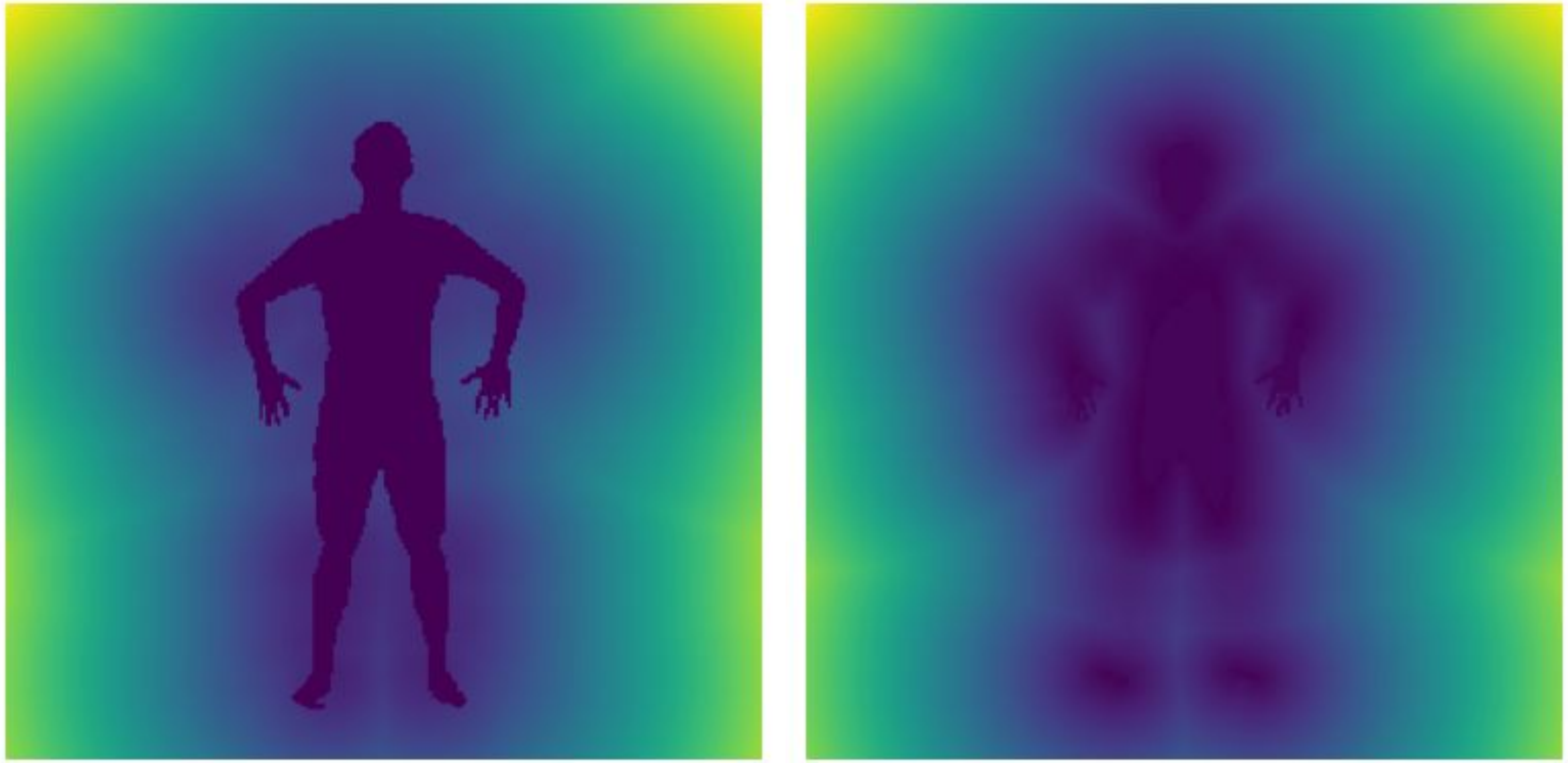


Figure 13: Monocular depth input represented as partial dicretized truncated SDF field. Left: Slice approximately around shape center. Right: Slice behind the shape, affected by occlusions.

Qualitative comparison

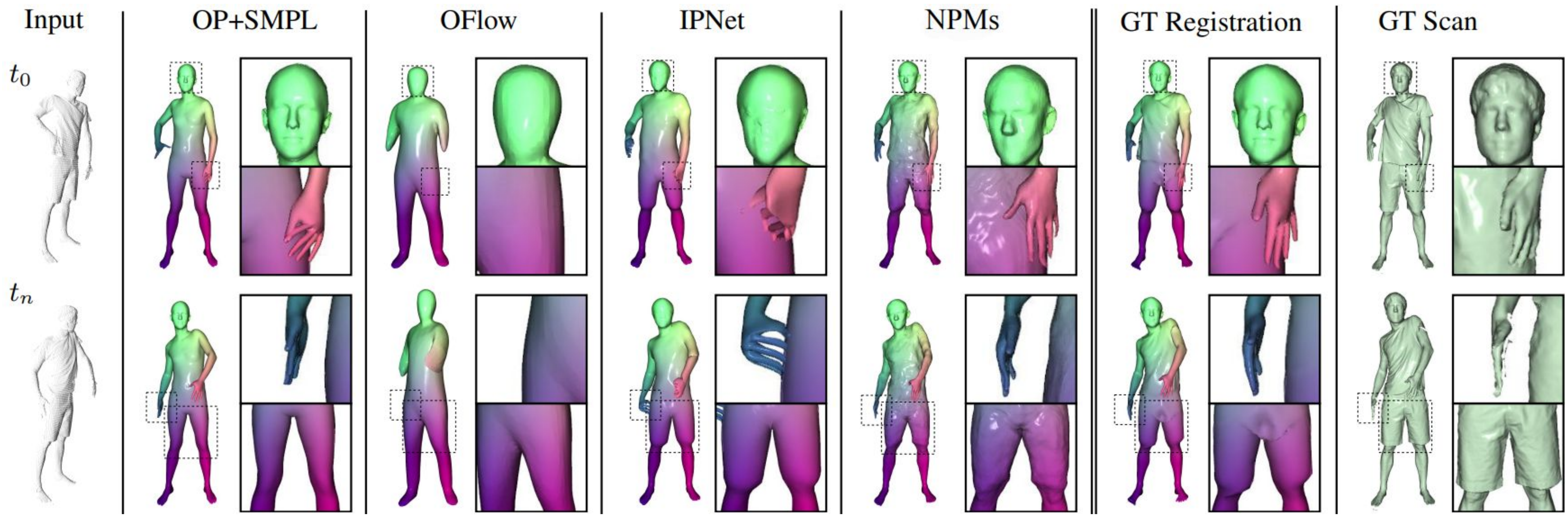


Figure 14: Qualitative results and comparison with state-of-the-art (at the time) models for non-rigid 4D reconstruction from monocular depth. Adapted from [6].

Quantitative comparison

Model	IoU	C-ℓ_2 $(\cdot 10^{-3})$	EPE $(\cdot 10^{-2})$
OpenPose+SMPL	0.68	0.243	2.82
OFlow	0.55	0.755	2.65
IP-Net	$\boxed{0.82}$	0.034	$\boxed{2.52}$
NPMs	$\boxed{0.83}$	0.022	$\boxed{0.74}$

Table 1: Quantitative results and comparison with state-of-the-art (at the time) models for non-rigid 4D reconstruction from monocular depth. Note that OFlow is evaluated on shorter sequences due to its limitations. Adapted from [6].

Details during interpolation

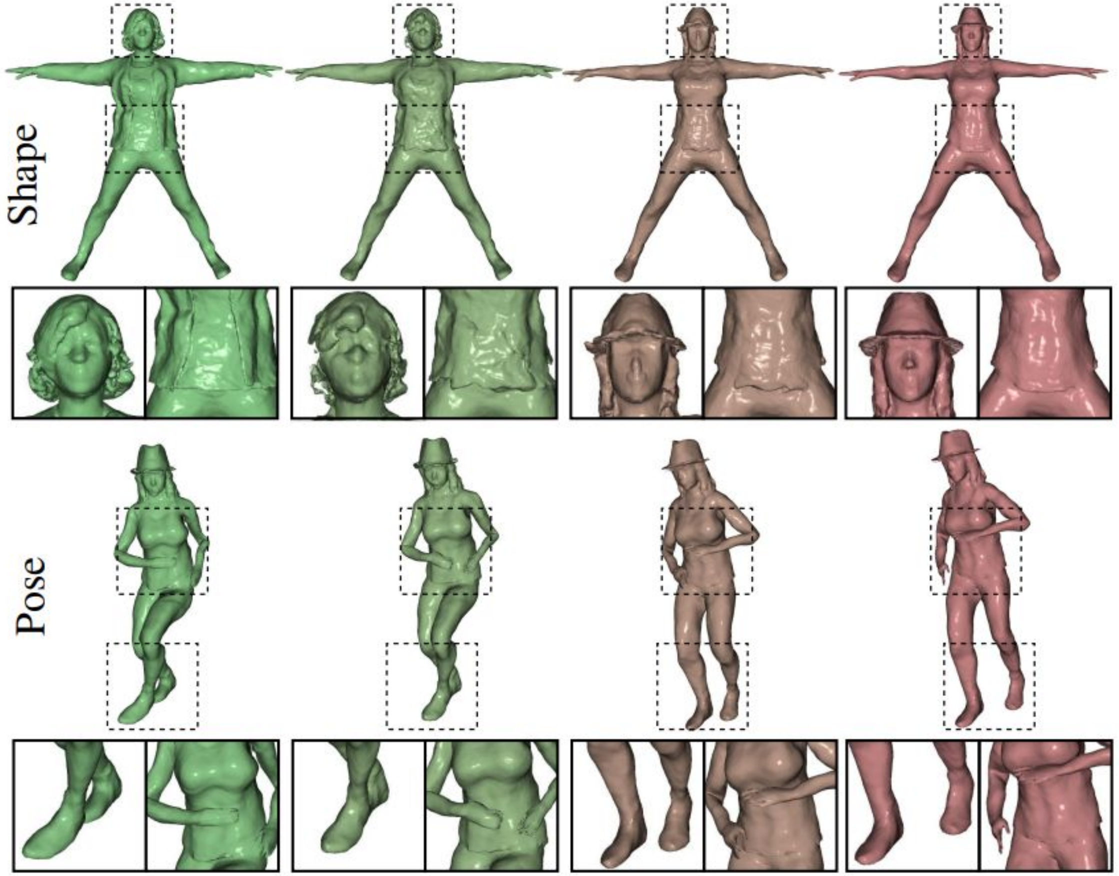


Figure 15: Shape and pose interpolation with NPMs. From [6]

Experiment replication

- Using a minimalistic test set (a single image)
- Observations:
  - Ambiguous results (shape code is inferred from entire sequence)
  - Computational cost prohibitive for longer sequences

Figure 16: Result of own minimalistic experiment.