Towards Efficient and Realistic Animation of 3D Garments with Deep Learning

Author: Hugo Bertiche
Supervisor: Sergio Escalera
Co-supervisor: Meysam Madadi

Thesis Defense

Contents

1. Chapter 1 - Introduction
2. Chapter 2 - SMPLR
3. Chapter 3 - CLOTH3D
4. Chapter 4 - DeePSD
5. Chapter 5 - PBNS
6. Chapter 6 - Neural Cloth Simulation
7. Chapter 7 - Conclusions

Chapter 1

Introduction

Cloth

- Exclusive of humans
- Feature of all societies
- Relevance:
  - Functionality
  - Status
  - Religion
  - Expression of self
  - ...
- Fashion Industry = $1.7 trillion

[1] Image source: https://en.wikipedia.org/wiki/Clothing

Motivation

- Entertainment industry:
  - TV & Films
  - Video-games
  - VR/AR
  - Social media content
  - ...
- Fashion industry:
  - Virtual try-on
  - e-commerce
  - Design
  - ...

[1] Image source: https://ardev.es/en/virtual-fitting-room-augmented-reality/
https://www.millenium.gg/noticias/34878.html

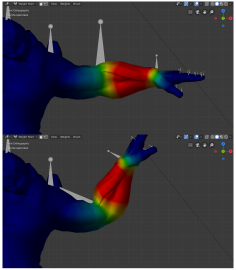
Motivation

- Classical solutions are too slow!
- Deep Learning fast inference time
- Deep Learning success in complex 3D tasks

Preamble

Preamble:
Linear Blend Skinning

Preamble: Linear Blend Skinning

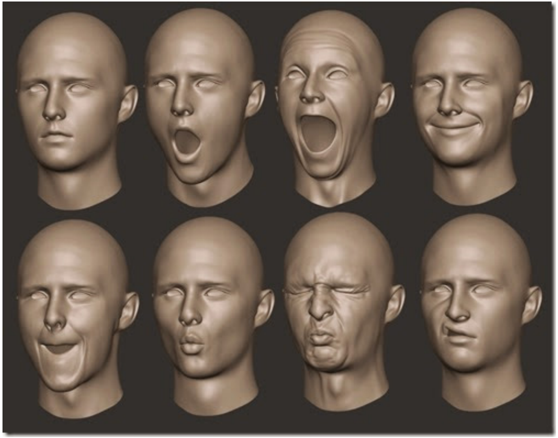


- Body mesh with N vertices
- Skeleton with K joints/bones
- Joint transforms $T \in \mathbb{R}^{K \times 4 \times 4}$
- Posed vertices: $v'_i = \sum_j^K w_{i,j} T_j v_i$

Preamble

Preamble:
Blend Shapes

Preamble: Blend Shapes



- M different shapes with N vertices
- Final shape $\mathbf{S}_f$:

$$\mathbf{S}_f = \sum_i^M w_i \mathbf{S}_i \tag{1}$$

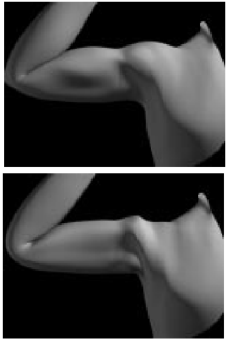
- Blend Shapes Matrix $\mathbf{S} \in \mathbb{R}^{M \times N \times 3}$
- Animate mesh $\leftrightarrow$ Animate $\mathbf{w} \in \mathbb{R}^M$

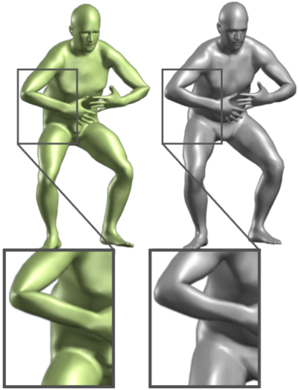
[1] Image source: `http://www.3dstudents.com/2015/02/blend-shapes.html`

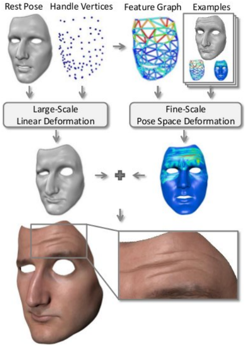
Preamble

Preamble:
Pose Space Deformation

Preamble: Pose Space Deformation







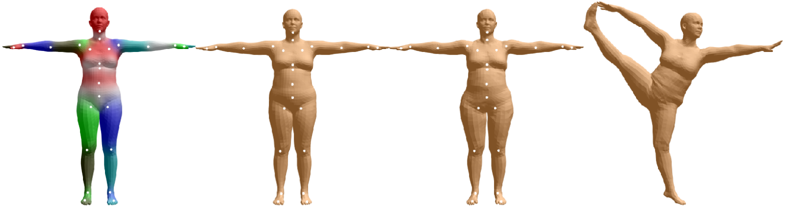
[1]

[1] Image sources: Lewis, J. P., Cordner, M., & Fong, N. (2000, July). Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 165-172).
Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. ACM transactions on graphics (TOG), 34(6), 1-16.
Bickel, B., Lang, M., Botsch, M., Otaduy, M. A., & Gross, M. H. (2008, July). Pose-Space Animation and Transfer of Facial Details. In Symposium on Computer Animation (pp. 57-66).

Preamble

Preamble:
SMPL

Preamble: What is SMPL?



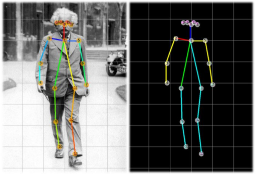
(a) $\bar{\mathbf{T}}, \mathcal{W}$ (b) $\bar{\mathbf{T}} + B_S(\vec{\beta}), J(\vec{\beta})$ (c) $T_P(\vec{\beta}, \vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$ (d) $W(T_P(\vec{\beta}, \vec{\theta}), J(\vec{\beta}), \vec{\theta}, \mathcal{W})$ [a]

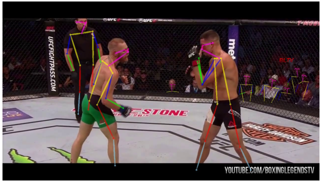
- $N = 6890$ vertices (> 20000 parameters)
- Pose $\theta \in \mathbb{R}^{72}$
- Shape $\beta \in \mathbb{R}^{10}$
- Gender $g \in \{0, 1\}$
- Total: 83 parameters

Chapter 2

SMPLR: Deep Learning Based SMPL Reverse for 3D Human Pose and Shape Recovery

Motivation





[1]

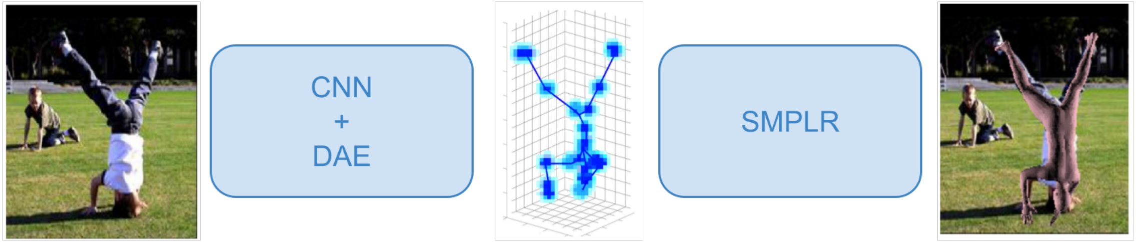
[1] https://towardsdatascience.com/realtime-multiple-person-2d-pose-estimation-using-tensorflow2-x-93e4c156d45f

Ning, G., Zhang, Z., & He, Z. (2017). Knowledge-guided deep fractal neural networks for human pose estimation.

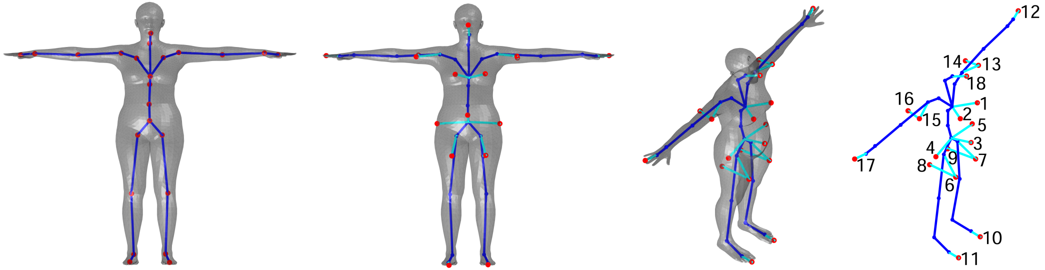
https://www.analyticsvidhya.com/blog/2018/06/ai-guardman-machine-learning-application-estimates-poses-detect-shoplifters/

Sasaki, K., Shiro, K., & Rekimoto, J. (2020, March). ExemPoser: Predicting Poses of Experts as Examples for Beginners in Climbing Using a Neural Network.

3D Humans from RGB: Overview



Joints and Landmarks



3D Human from RGB: SMPL Reverse

Normalized joints:

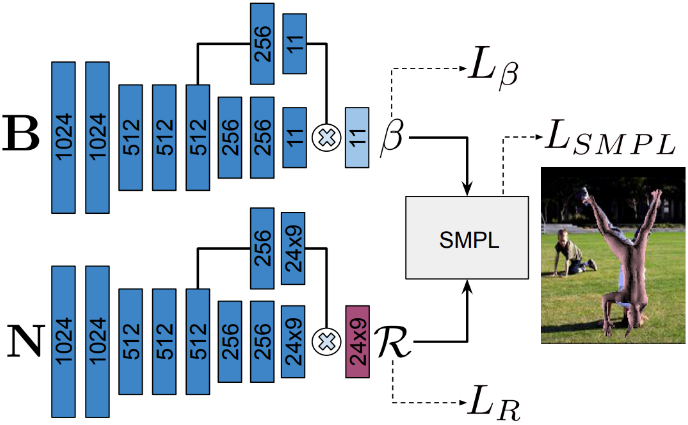
$$\mathbf{N}_i = \frac{\mathbf{J}_i - \mathbf{J}_{k(i)}}{\|\mathbf{J}_i - \mathbf{J}_{k(i)}\|_2} \qquad (2)$$

Normalized bones:

$$\mathbf{B}_i^* = \|\mathbf{J}_i^* - \mathbf{J}_{k(i)}^*\|_2, \qquad (3)$$

$$\mathbf{B}_i = \|\mathbf{J}_i - \mathbf{J}_{k(i)}\|_2 - \mathbf{B}_i^* \qquad (4)$$

Losses $\mathcal{L}_* := \mathrm{L1}$



Results

Conclusions

SMPLR contributions:

- 3D Humans from still RGB images
- SMPLR: recovers SMPL parameters
- DAE: recovers from structured error and 2D/3D lifting
- Joints and Landmarks
- State-of-the-art results

Chapter 3

CLOTH3D: Clothed 3D Humans

Motivation

- Data-hungry approaches
- No large scale dataset
- Push research

Motivation

Dataset	3DPW[18]	BUFF[35]	Untitled[29]	3DPeople[24]	TailorNet[21]	CLOTH3D
Resolution	2.5cm	0.4cm	1cm	-[1]	1cm	1cm
Missing	x	✓	x	x	x	x
Dynamics	x	✓	x	x	x	✓
Garments	18[2]	10∼ 20	3[3]	High[4]	20	11.3K
Fabrics	x	x	x	x	x	✓
Poses[5]	Low	Low	Very low	Low	1782	High
Subjects	18[2]	6	2K	80	9	8.5K
Layered	x	x	✓	-[1]	✓	✓
#samples	51k	11K	24K	2.5M	55.8k	2.1M
Type	Real	Real	Synth.	Synth.	Synth.	Synth.
RGB	✓	x	✓	✓	x	x
GT error	26mm	1.5-3mm	None	None	None	None

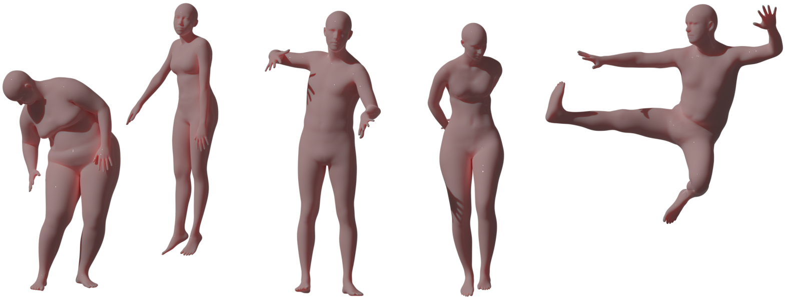
[18] 3DPW: 3D Poses in the wild dataset
[35] BUFF: Bodies Under Flowing Fashion, 4D dataset
[29] Learning a shared shape space for multimodal garment design
[24] 3DPeople: Modeling the Geometry of Dressed Humans
[21] TailorNet: Predicting Clothing in 3D as a function of Human Pose, Shape and Garment Style

Data Generation: Overview

Consistent garment procedural generation

1. Pick template
2. Change shape
3. Cut
4. Resize
5. Combine (optional)

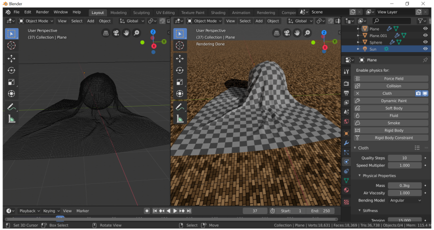
Data Generation: Humans

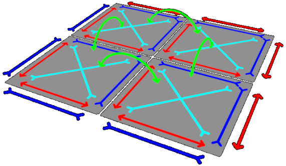


Walk	Animal	Fight	Jump	Run	Sing	Wait	Swim	Story	Sports	Dance	Yoga	Spin
27.49%	10.79%	4.38%	2.78%	2.49%	2.38%	2.31%	1.97%	1.70%	1.63%	1.37%	1.01%	0.90%

Exercise	Climb	Carry	Stand	Wash	Balancing	Trick	Sit	Interact	Drink	Pose	Others
0.84%	0.71%	0.67%	0.66%	0.63%	0.54%	0.51%	0.28%	0.20%	0.14%	0.14%	33.48%

Data Generation: Simulation



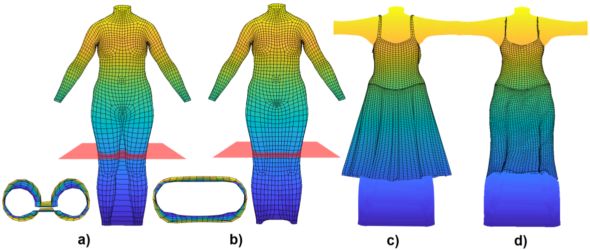


Dataset: samples

Baseline

Baseline

Baseline: Pre-process



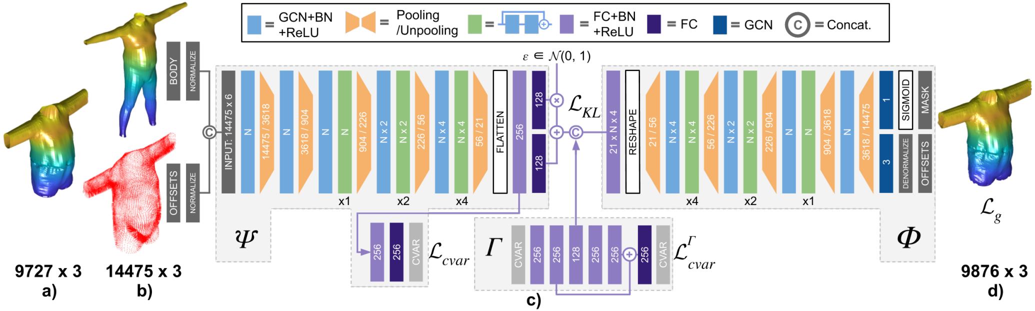
- Garments vs Body: NR-ICP
- Nearest Neighbour matching
- Compute offsets
- Mask: Garment topology
- * SuperSMPL
- * SMPL-Skirt topology

Heterogeneus garments V into uniform space[2]:

$$\mathbb{R}^{|V|\times 3} \rightarrow \mathbb{R}^{14475\times 4} \tag{5}$$

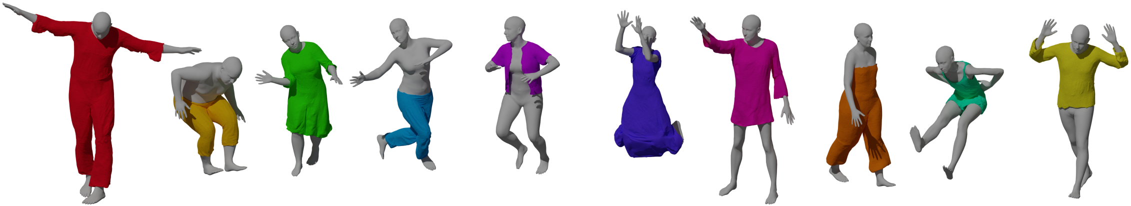
[2] 4th dimension for body mask.

Baseline: GraphCVAE

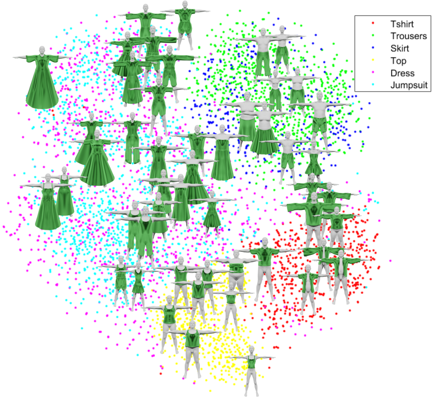


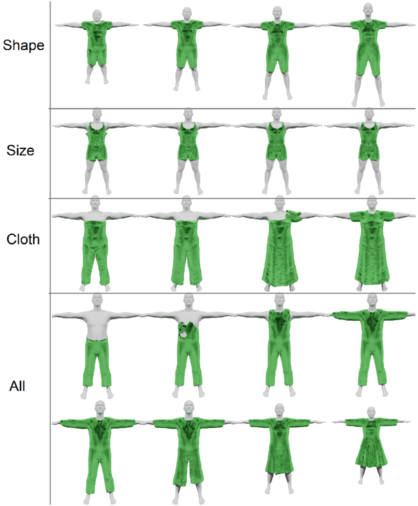
- Static vs. Dynamic: disentangled
- Latent space for garment topology
- Latent space for dynamics (wrinkles)

Baseline: results



Baseline: Static latent space





Conclusion

CLOTH3D contributions:

- First large scale 3D dataset on clothes
- Huge topology, shape and dynamics variability
- Aid researchers on current challenges:
    - Heterogeneus meshes
    - High and low frequency information
    - Complex dynamics
    - Collisions
    - ...
- Baseline inspired in SOTA:
    - Learns continuous topologic space for later use in generation
    - Learns garment dynamics as a function of pose and garment style (topology, shape and tightness)

Chapter 4

DeePSD: Automatic Deep Skinning and Pose Space Deformation for 3D Garment Animation

Motivation

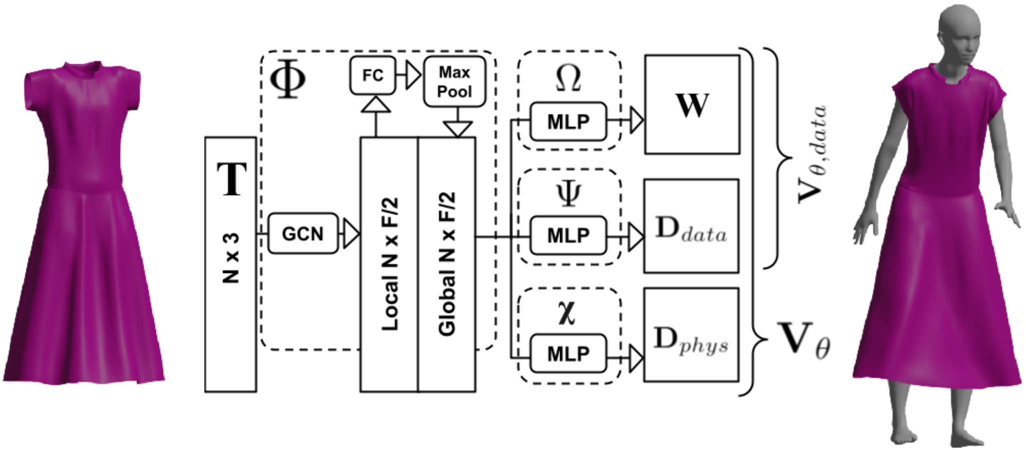
- Outfit variability is limitless
- Discard outfit generation
- Focus on outfit animation

$$f : \theta \rightarrow \mathbf{V}_\theta \tag{6}$$

Contributions

- Outfit generalization
- Compatibility
- Physically consistency
- Intuitive workflow

Architecture



Training

Supervised loss:

$$\mathcal{L}_{\text{data}} = \sum \left\| \mathbf{V}_{\theta,\text{data}} - \mathbf{V}_{\text{PBS}} \right\|^2 \tag{7}$$

Unsupervised loss:

$$\begin{aligned}
\mathcal{L}_{\text{cloth}} &= \mathcal{L}_E + \lambda_B \mathcal{L}_B = \sum_{e \in E} \| e - e_{\mathbf{T}} \| + \lambda_B \Delta(\mathbf{n})^2 \\
\mathcal{L}_{\text{collision}} &= \sum_{(i,j) \in A} \min(\mathbf{d}_{j,i} \cdot \mathbf{n}_j - \epsilon, 0)^2 \\
\mathcal{L}_{\text{phys}} &= \mathcal{L}_{\text{cloth}} + \lambda_{\text{collision}} \mathcal{L}_{\text{collision}}
\end{aligned} \tag{8}$$

vs CLOTH3D Baseline

vs TailorNet

DeePSD TailorNet DeePSD TailorNet DeePSD TailorNet

Virtual Try-on

Conclusions

- Outfit generalization
  - Scalability
  - Intuitive workflow for artists
- Compatibility
  - Low engineering effort to adapt
  - Efficient!
- Physically based loss
  - Pose augmentation (train)
  - Robust predictions (test)

Chapter 5

PBNS: Physically Based Neural Simulation for Unsupervised Garment Pose Space Deformation

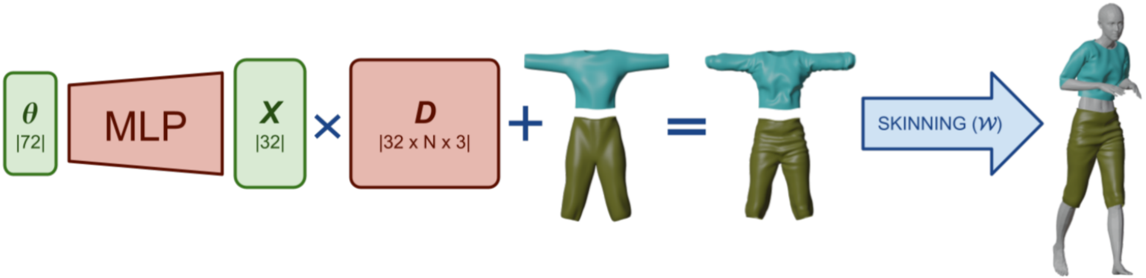
Motivation

- High applicability of outfit animation
- Deep learning success in 3D
- Deep learning inference efficiency
- Data hungry algorithms

Contributions

- Unsupervised
- Physically consistent results
- Cloth-to-cloth interaction
- Extremely efficient

Architecture



- Outfit skinned w.r.t. body skeleton
- θ processed by MLP
- Pose Space Deformations as Matrix D
- MLP and D are learnt

Loss

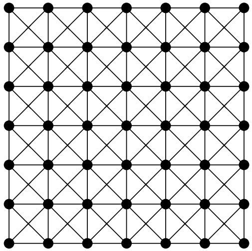
Energy based formulation:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{cloth} + \mathcal{L}_{collision} + \mathcal{L}_{gravity} + \mathcal{L}_{pin} \\ \mathcal{L}_{cloth} &= \mathcal{L}_{edge} + \mathcal{L}_{bend} \end{aligned} \tag{9}$$

- Internal energies $\rightarrow$ Cloth
- External energies $\rightarrow$ Collision, Gravity and Pin

Loss - Edge

Inspired in mass-spring models



Edge loss:

$$\mathcal{L}_{edge} = \lambda_e \|E - E_T\|^2 \qquad (10)$$

- E : predicted edge lengths
- E_T : template edge lengths
- λ_e : balancing weight

Spring energy:

$$E_{spring} = \frac{1}{2} k \Delta x^2 \qquad (11)$$

Loss - Bend

Bend loss:

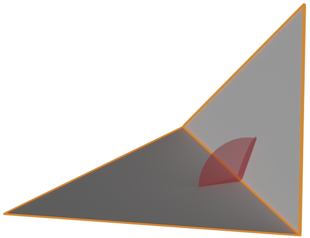
$$\mathcal{L}_{bend} = \lambda_b \Delta(\mathbf{N})^2 \tag{12}$$

Squared discrete Laplacian for graph:

$$(\Delta\phi)(v) = \sum_{w:\mathcal{N}(v)} [\phi(v) - \phi(w)]^2 \tag{13}$$

- Graph $G = (V, E)$:
  - V : faces
  - E : face adjacency
- $\phi(\cdot)$ computes face normal
- λ_b : balancing weight

Hinge-like energy



Penalizes adjacent face normals differences

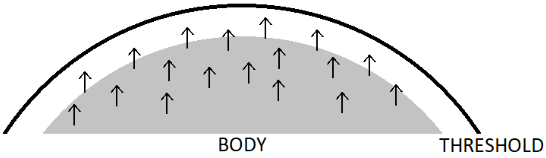
Loss - Collision

Collision loss:

$$\mathcal{L}_{collision} = \lambda_c \sum_{(i,j)\in A} min(\mathbf{d}_{j,i} \cdot \mathbf{n}_j - \epsilon, 0)^2 \tag{14}$$

- A : cloth-to-body correspondences
- $d_{j,i}$: body-to-cloth vector
- n_j : j-th body vertex normal
- ϵ : threshold
- λ_c : balancing weight

Repulsion gradients (forces)



Loss - Collision - Cloth-to-cloth

Cloth-to-cloth interaction:

- Multiple ordered layers of cloth
- Apply $\mathcal{L}_{collision}$ iteratively
- For layer n, consider layers $1, ..., n-1$ as body
- Gradients push both layer vertices

Loss - Gravity / Pin

Gravity loss:

$$\mathcal{L}_{gravity} = \sum m \cdot g \cdot h \qquad (15)$$

- m : vertex mass
- g : gravity
- h : height

Pin loss:

$$\mathcal{L}_{pin} = \lambda_{pin} \sum_i b_i dt_i^2 \qquad (16)$$

- b_i : i-th vertex is pinned $\{0, 1\}$
- dt : i-th vertex deformation
- λ_{pin} : balancing weight

Results

Results - Performance

	Single	Batch
CPU	213 FPS	1235 FPS
GPU	455 FPS	14286 FPS

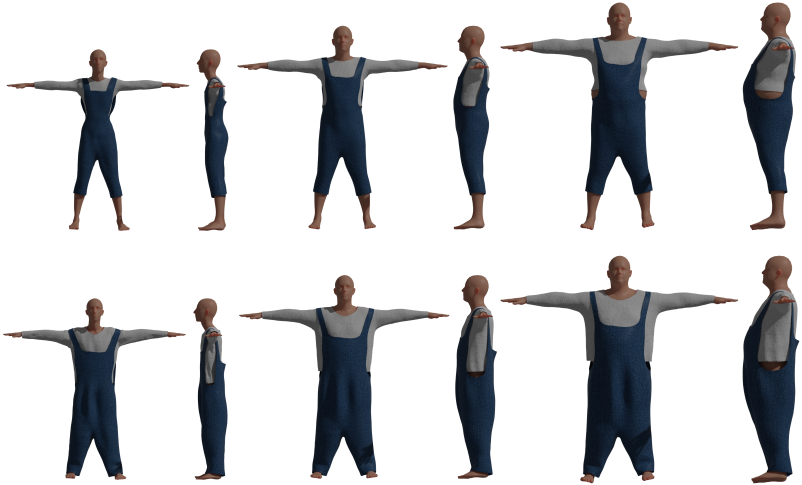
- Run on GTX1080Ti
- Outfit with 23.7k triangles
- Training: 1-2 min/epoch
- Memory: few MBs

Results - Layers

Results - Custom Avatars[3]

[3]https://www.mixamo.com/

Results - Resizing



Conclusion

PBNS properties:

- Unsupervised
- Physical consistency
- Control cloth properties
- Multiple layers of cloth
- Boots, gloves, ...
- Extremely efficient: +14K samples/s
- Simple formulation: minimal engineering effort

Chapter 6

Neural Cloth Simulation

Motivation

Contributions

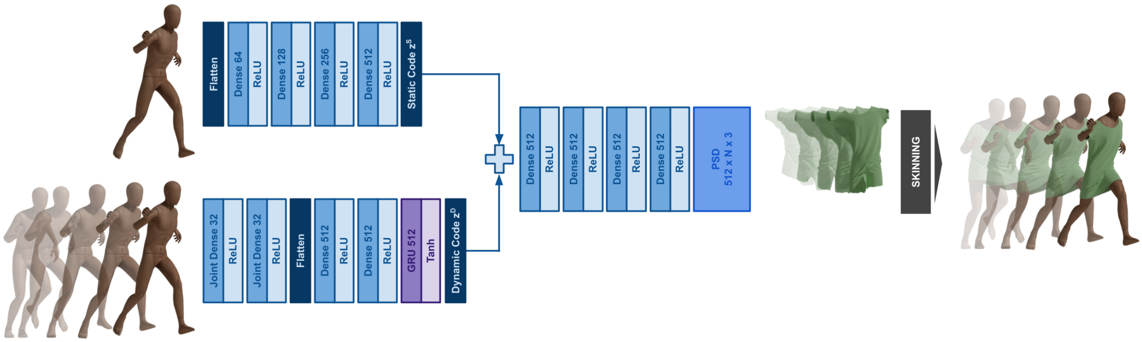
- Unsupervised Cloth Dynamics
- Disentangled Cloth Subspace
- Domain Analysis

Static vs Dynamic

Static

Dynamic

Disentangled Network



Losses

Loss

$$\mathcal{L} = \mathcal{L}_{\text{cloth}} + \mathcal{L}_{\text{bending}} + L_{\text{collision}} + \mathcal{L}_{\text{inertia}} + \mathcal{L}_{\text{gravity}} = E(x;\theta) \tag{17}$$

- Loss = Energy
- Training = Simulating

Cloth Model

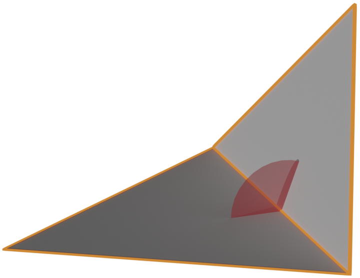
- Mass-spring
- Finite Element Methods
- Saint-Venant Kirchhoff
- ...

Bending

Bending Loss

$$\mathcal{L}_{\text{bending}} = k_b \frac{l^2}{8a}(\phi_t - \phi^R)^2 \qquad (18)$$

- k_b : bending stiffness
- l : edge length
- a : triangle area
- ϕ_t : dihedral angle at t
- ϕ^R : rest dihedral angle

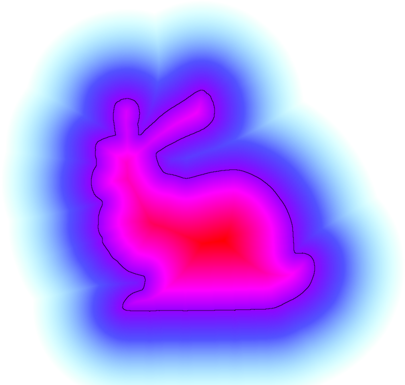


Collisions

Collision Loss

$$\mathcal{L}_{\text{collision}} = k_c \text{min}(\text{d}(x_t; \theta_t) - \epsilon, 0)^2, \tag{19}$$

- k_c : collision *stiffness*
- $d(\cdot)$: signed distance
- x_t : cloth vertex locations at t
- θ_t : body parameterization
- ϵ : small threshold (robustness)



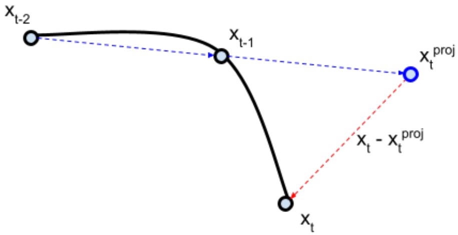
Inertia

Inertia Loss

$$\mathcal{L}_{\text{inertia}} = \frac{1}{2\Delta t^2} m(x_t - x_t^{\text{proj}})^2$$
$$x_t^{\text{proj}} = 2x_{t-1} - x_{t-2} \tag{20}$$

- Δt : time step
- m : vertex mass
- x_t : cloth vertex locations at t

Do NOT back-propagate through x_{t-1} and x_{t-2}!!!



Gravity

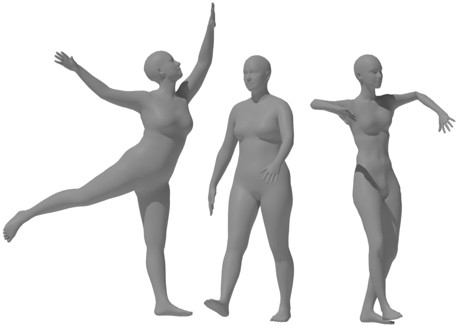
Gravity Loss

$$\mathcal{L}_{\text{gravity}} = -M x_t g \tag{21}$$

- M : mass matrix
- x_t : vertex locations at t
- g : gravity

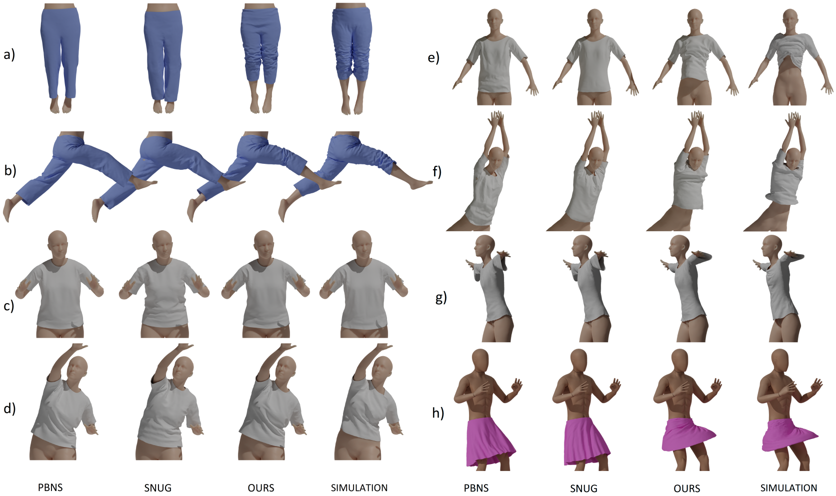
Data: Pose Sequences

Mixamo

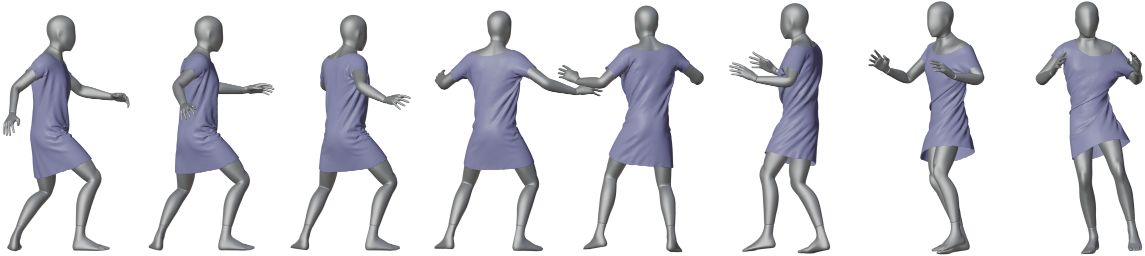


SMPL

PBNS vs SNUG vs NCS vs Simulation



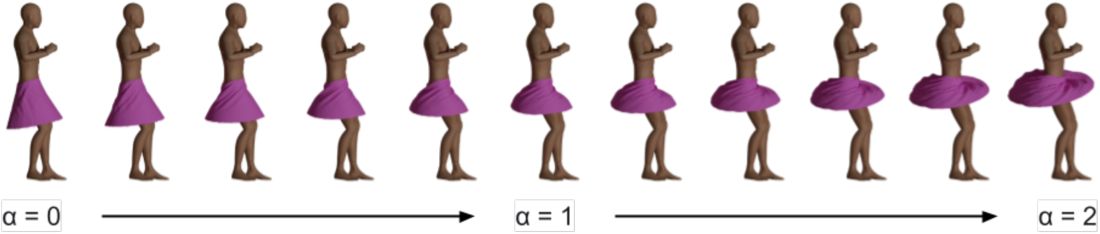
Qualitative



Motion Control

$$z = z^S + z^D \tag{22}$$

$$z_\alpha = z^S + \alpha z^D \tag{23}$$



Conclusions

- Unsupervised Cloth Dynamics
- Novel set of descriptors
- Disentangled subspaces for Static/Dynamic cloth deformations
  - Improved generalization
  - Novel motion augmentation (train)
  - Novel motion control (test)
- Compatible with arbitrary 3D characters/garments
- Real-time performance
- In-depth domain analysis

Chapter 7

Conclusions

Achieved Milestones

- SMPLR: retrieve 3D humans from RGB images
- CLOTH3D: First large scale clothing dataset
- CLOTH3D Baseline: 3D generative model and learnt space for all garments
- DeePSD: automatic animation of arbitrary outfits
- PBNS: first unsupervised learning-based solution
- Neural Cloth Simulation: unsupervised cloth dynamics

What's next?

- Faster convergence & more efficient methods
- Garment & Body generalization
- Universal deep cloth simulator
- More...

Thesis Publications

- Meysam Madadi et al. (2020). "SMPLR: Deep learning based SMPL reverse for 3D human pose and shape recovery". In: *Pattern Recognition*, p. 107472
- Hugo Bertiche et al. (2020). "CLOTH3D: Clothed 3D Humans". In: *European Conference on Computer Vision*. Springer, pp. 344–359
- Meysam Madadi et al. (2021a). "Deep unsupervised 3D human body reconstruction from a sparse set of landmarks". In: *International Journal of Computer Vision* 129.8, pp. 2499–2512
- Meysam Madadi et al. (2021b). "Learning Cloth Dynamics: 3D+Texture Garment Reconstruction Benchmark". In: *Proceedings of the NeurIPS 2020 Competition and Demonstration Track*. Ed. by Hugo Jair Escalante et al. Vol. 133. Proceedings of Machine Learning Research. PMLR, pp. 57–76. URL: https://proceedings.mlr.press/v133/madadi21a.html
- Hugo Bertiche et al. (2021a). "Deep Parametric Surfaces for 3D Outfit Reconstruction from Single View Image". In: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pp. 1–8. DOI: 10.1109/FG52635.2021.9667017
- Hugo Bertiche et al. (2021b). "Neural Implicit Surfaces for Efficient and Accurate Collisions in Physically Based Simulations". In: *CoRR* abs/2110.01614. arXiv: 2110.01614. URL: https://arxiv.org/abs/2110.01614
- Hugo Bertiche et al. (2021d). "DeePSD: Automatic deep skinning and pose space deformation for 3D garment animation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5471–5480
- Hugo Bertiche et al. (2021c). "PBNS: Physically Based Neural Simulation for Unsupervised Garment Pose Space Deformation". In: *ACM Trans. Graph.* 40.6. ISSN: 0730-0301. DOI: 10.1145/3478513.3480479. URL: https://doi.org/10.1145/3478513.3480479