Virtual Humans – Winter 23/24

Lecture 10_2 – Humans and NeRF

Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics
Novel View Synthesis for Humans

Task: Novel view synthesis from a sparse multi-view video

4-view video

Novel view synthesis of dynamic human (Our result)
Human models using NeRF

Challenge: It is ill-posed to learn 3D representations from very sparse observations

Four input images  Novel view synthesis by NeRF [3]

Neural Body: Implicit Neural Representations with Structured Latent Codes for NVS of Dynamic Humans
Neural Body: Key Idea

\[ \mathcal{Z} = \{ z_1, z_2, \ldots, z_{6890} \} \]
Neural Body: Key Idea

Structured latent codes

\[ \mathbf{Z} = \{ z_1, z_2, \ldots, z_{6890} \} \]

Sparse latent code, not defined for all points in 3D

Peng et al. CVPR 21
Neural Body: Method

(a) Code diffusion
Neural Body: Method

(a) Code diffusion

Structured latent codes $\mathcal{E}$

Human pose $S_t$

SparseConvNet

Latent code volume

Latent code $\psi(x, \mathcal{E}, S_t)$

Peng et al. CVPR 21
Neural Body: Method

(a) Code diffusion

(b) Density and color regression

Peng et al. CVPR 21
Neural Body: Results

Input Views  NeuralBody  3D reconstruction

Peng et al. CVPR 21
Neural Body: Conclusion

• Use SMPL mesh as structure:
  + Strong human prior and preserves human shape.

Peng et al. CVPR 21
Neural Body: Conclusion

• Use SMPL mesh as structure:
  + Strong human prior and preserves human shape.
  - Introduces artifacts in clothing and complex motions that are not captured by the SMPL model.

Peng et al. CVPR 21
Neural Body: Conclusion

• Use SMPL mesh as structure:
  + Strong human prior and preserves human shape.
  - Introduces artifacts in clothing and complex motions that are not captured by the SMPL model.
  - Only works for multi-view setup.

Peng et al. CVPR 21
HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video

- Given a monocular video (a) of a human performing complex movement, e.g., dancing (left), HumanNeRF creates a free-viewpoint rendering for any frame in the sequence (b).
- Deformation and skinning formulation similar to NeuralGIF

Weng et al. CVPR 22
HumanNeRF: Key Idea

• Split the deformation into:
  1. Human articulation
  2. Non-rigid pose dependent deformation

Similar to NerualGIF

Weng et al. CVPR 22
HumanNeRF: Key Idea

• Split the deformation into:
  1. Human articulation
  2. Non-rigid pose dependent deformation

• Skinning weights using forward skinning.

Similar to NerualGIF

Similar to SNARF

Weng et al. CVPR 22
HumanNeRF: Method

$T_{\text{ske}}(\mathbf{x}, \mathbf{p}) = \sum_{i=1}^{K} w^i_o(\mathbf{x})(R_i \mathbf{x} + t_i)$,  

Observed to canonical

Skinning weight obtained using forward skinning

Weng et al. CVPR 22
HumanNeRF: Method

\[ T_{NR}(x, p) = MLP_{\theta_{NR}}(\gamma(x); \Omega), \]

- **Positional encoding**
- **Set of joint rotations**

Weng et al. CVPR 22
HumanNeRF: Method

Weng et al. CVPR 22
HumanNeRF: Method
More on Human and NeRFs

• Animatable NeRF, Peng et al., ICCV2021
• H-NeRF, Xu et al., NeurIPS 2021
• NeuMan, Jian et al., ECCV 2022
• DoubleField, Shao et al., CVPR 2022

• ....... And many more.
Limitation of NeRF/Implicit Representations

• 4. Expensive training:
  • Training is slow (10 hours-up to few days)
  • Inference is also not real time
Mixture of Volumetric Primitives

• Combines the advantages of volumetric and primitive-based approaches for:
  • High performance decoding
  • Efficient rendering

• A novel motion model for voxel grids for scene motion, minimization of primitive overlap to increase the representational power.
Mixture of Volumetric Primitives: Key Idea

• Combining primitives and volumetric representation:

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Key Idea

• Combining primitives and volumetric representation:

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Key Idea

- Combining primitives and volumetric representation:

\[ V_k = \{ t_k, R_k, s_k, V_k \} \]

Translation, rotation and scale of grid/primitives

\[ V_k \in \mathbb{R}^{4 \times M_x \times M_y \times M_z} \]

Where each feature grid contains color and density and M is grid resolution

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Key Idea

• Combining primitives and volumetric representation:
Mixture of Volumetric Primitives: Method
Mixture of Volumetric Primitives: Method

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Method

\[ V_k \in \mathbb{R}^{4 \times M_x \times M_y \times M_z} \]

Where each feature grid contains color and density and \( M \) is grid resolution.

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Method

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Method

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Method

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Method

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Training

\[ \mathcal{L}(\Theta; I_p) = \mathcal{L}_{\text{pho}}(\Theta; I_p) + \mathcal{L}_{\text{geo}}(\Theta) + \mathcal{L}_{\text{vol}}(\Theta) + \mathcal{L}_{\text{del}}(\Theta) + \mathcal{L}_{\text{kld}}(\Theta) \]

\[ \mathcal{L}_{\text{pho}} = \lambda_{\text{pho}} \frac{1}{N_{\varphi}} \sum_{p \in \mathcal{P}} \| I_p - \bar{I}_p(\Theta) \|^2 \]

Difference between predicted and GT image

\[ \mathcal{L}_{\text{vol}} = \lambda_{\text{vol}} \sum_{i=1}^{N_{\text{prim}}} \text{Prod}(s_i) \]

Volumetric primitive to be as small as possible

\[ \mathcal{L}_{\text{geo}} = \lambda_{\text{geo}} \frac{1}{N_{\text{mesh}}} \sum_{i=0}^{N_{\text{mesh}}} \| \mathbf{v}_i - \bar{\mathbf{v}}_i(\Theta) \|^2 \]

Difference regressed vertex position and GT vertex

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Results

(a) GT  (b) NV  (c) 8 prim.  (d) 512 prim.  (e) 32k prim.  (f) 256 prim.*

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives: Results

A stronger primitive volume prior leads to less overlap and thus speeds up raymarching

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

How much the result depends on initialization (or the guide coarse mesh)?
Mixture of Volumetric Primitives

How much the result depends on initialization (or the guide coarse mesh)?

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

How much the result depends on initialization (or the guide coarse mesh)?

- **Alpha Fade**: Windowing function adds an inductive bias to explain the scene’s contents via motion instead of volumetric opacity.

\[
W(x, y, z) = \exp\left(-\alpha(x^\beta + y^\beta + z^\beta)\right)
\]

\[
W(x, y, z) \in \mathbb{R}^M^3
\]

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

How much the result depends on initialization (or the guide coarse mesh)?

(a) GT  (b) No alpha fade  (c) With alpha fade

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

+ Combine volumetric and primitive based approach for generizable representation of dynamic scenes.
  • Fast to render
  • Represent translucent parts, thing structures

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

Combine volumetric and primitive based approach for generizable representation of dynamic scenes.
  • Fast to render
  • Represent translucent parts, thing structures

Strong prior about structure of underlying shape via coarse shape and derived primitives

Lombardi et al. SIGGRAPH 21
Mixture of Volumetric Primitives

Combine volumetric and primitive based approach for generizable representation of dynamic scenes.
- Fast to render
- Represent translucent parts, thing structures

Strong prior about structure of underlying shape via coarse shape and derived primitives

No prior/information about the structure of motion of underlying object.
- Difficulty to model human motion

Lombardi et al. SIGGRAPH 21
Drivable Volumetric Avatars

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars

**Key Idea: Articulated Primitives**
- Use SMPL posed mesh, $M_\theta$ as a guide mesh and define primitives using it.
- Initialise primitive by uniformly sampling UV space and mapping each primitive to closest texel $t_k(\theta)$.
- Transform primitives using transformation matrices of SMPL joints and skinning weights.

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars: Method

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars: Method

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars: Method

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars: Method

Remelli et al. SIGGRAPH 22
Drivable Volumetric Avatars: Results

FBCA  NeuralBody  OURS  Ground Truth

Remelli et al. SIGGRAPH 22
Conclusion:

• For NVS of humans, it helps to introduce human shape and structure prior in the method.
  • Provides controlability
  • Preserves human shape/structure

• Mixture of volumetric primitives helps:
  • Efficient rendering
  • Preserving fine details and model translucent structures
More on Human and NeRFs

• Animatable NeRF, Peng et al., ICCV 2021
• H-NeRF, Xu et al., NeurIPS 2021
• NeuMan, Jian et al., ECCV 2022
• DoubleField, Shao et al., CVPR 2022
•
• ....... And many more