BAIR Commons Year 2 Update
lbsNeRF: Animatable Volumetric Avatars from Videos

Triad: Hang Gao, Shubham Tulsiani, Angjoo Kanazawa
{hangg,kanazawa}@berkeley.edu, shubhtuls@fb.com

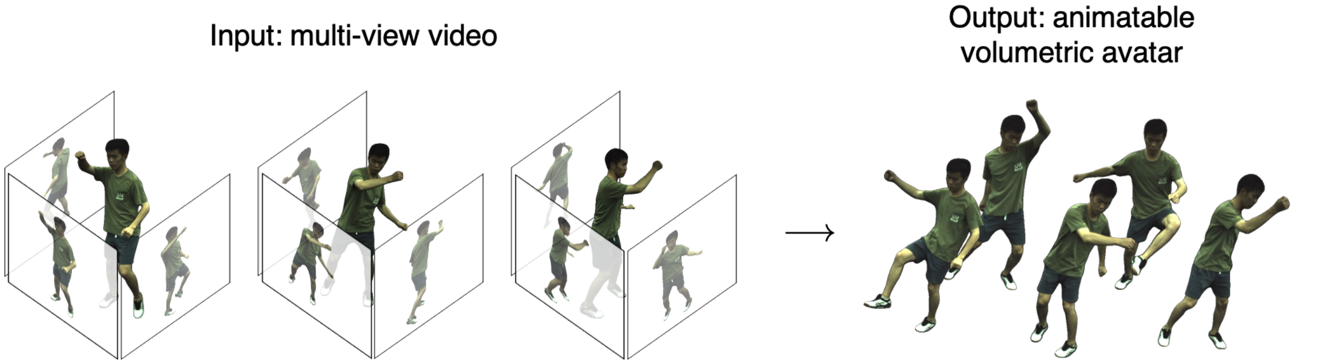

Figure 1: **Animatable Volumetric Avatars from videos.** We aim for a computational framework that can recover animatable avatars from multi-view videos, encoded as implicit 3D neural volumes. Such avatars are animatable, i.e. repose-able given any novel body-pose parameters.

1. Introduction and Related Works

Neural radiance fields (NeRF) [1] have emerged as a promising representation for encoding geometry and appearance of static scenes and objects. However, extending these representations to capture the non-rigid deformations common in categories such as humans and animals remains an open challenge. Towards this goal, we present lbsNeRF, a framework that learns an actor-specific neural avatar from multi-view videos with associated skeleton motions, and subsequently allows rendering under arbitrary query articulation and viewpoints. Inspired by classic works on skinning models that allow controlled deformation of meshes [2], our approach similarly constrains the allowed deformation under articulation. Our approach models appearance under articulation using a canonical space NeRF which is associated to the view space via a (neural) blending weight field that induces a per-point transformation.

Unlike existing works [3, 4] that relies on predefined surface models e.g. SMPL [5], for the first time, we show that lbsNeRF allows learning using only posed skeleton, and that this helps scale our approach to other generic categories e.g. cats and elephants. We empirically validate our approach across multiple datasets for both humans *and* animals. We show that our method allows generalization to unseen poses, while also performing comparably to prior methods which require stronger supervision.

2. Novelty and Innovation

For simplicity, in this section we only discuss our core contribution. Our network structure is the same as [4], consisting of a LBS weight network and a neural radiance field network.

Precisely, our main novelty is a chain of regularization that enforces rigidity of the 3D body transform as well as meaningful LBS weights that respect kinematic structure of skeletons. To be specific, we has a bone regularization that enforces the LBS weight network to predict the weights that correspond to parents for every point sample along the skeletal bones. This simple technique prevents early collapse of the training and ensures basic kinematic consistency. We have also developed a novel as-rigid-as-possible (ARAP) regularization for meaningful articulation. For each sample for the weight network, we assign a soft pseudo-label to its nearest joint. Effectively, this encodes the rigidity prior into our network since LBS weights have strong correlation to the distance between a surface point and candidate bones [6]. In our case, we generalize it into a entire volume with free spaces.

3. Current State and Future Plan

Our project is still on-going and we hope to share more details when finished.

We have demonstrated on existing benchmarks as used by [4] and find our method compare favorably, even without using any surface supervision.

We have tried our approach to other synthetic animal animation sequences we extracted from Blender but found that the sequences are rather noisy and our method does not work as good as on human data. Even in this case, our preliminary results show that our method performs comparably with existing works that requires ground-truth surface to start with.

To fully demonstrate the strength of our approach, we plan on next working on real animal captures where no surface parametric model is available.

References

[1] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1

[2] Thomas J Cashman and Andrew W Fitzgibbon. What shape are dolphins? building 3d morphable models from 2d images. *TPAMI*, 2012. 1

[3] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *CVPR*, 2021. 1

[4] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Animatable neural radiance fields for human body modeling. *arXiv preprint arXiv:2105.02872*, 2021. 1, 2

[5] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multiperson linear model. *SIGGRAPH*, 2015. 1

[6] Ilya Baran and Jovan Popović. Automatic rigging and animation of 3d characters. *SIGGRAPH*, 2007. 1