Sparse Motion Semantics for Contact-Aware Retargeting

Théo Cheynel
Kinetix
LIX, École Polytechnique, CNRS, IP Paris

Thomas Rossi
Kinetix

Baptiste Bellot-Gurlet
Kinetix

Damien Rohmer
LIX, École Polytechnique, CNRS, IP Paris

Marie-Paule Cani
LIX, École Polytechnique, CNRS, IP Paris

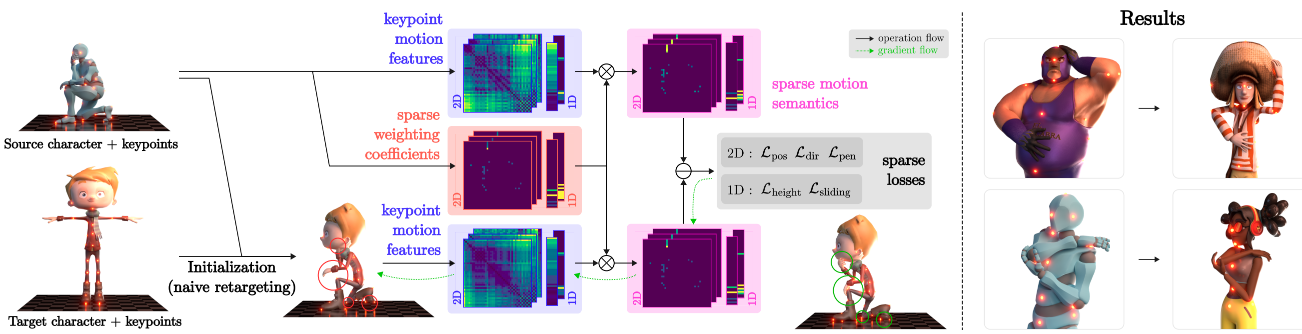

Figure 1: Overview of our retargeting method and results

ABSTRACT

This paper presents a method for retargeting motion onto a character with a completely different skeleton and mesh. We simplify the source and target meshes into a short list of corresponding keypoints, used to define a sparse motion semantic representation. We reformulate the retargeting problem as the minimization of new objective functions derived from this semantic representation, making it possible to optimize the target motion to achieve a contact-aware retargeting in real-time.

ACM Reference Format:

Théo Cheynel, Thomas Rossi, Baptiste Bellot-Gurlet, Damien Rohmer, and Marie - Paule Cani. 2023. Sparse Motion Semantics for Contact-Aware Retargeting. In *Proceedings of MIG '23*. ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION AND RELATED WORK

Transferring the motion of a character onto another one, called motion retargeting, is a crucial aspect of character animation. Complexity comes from the wide variety of 3D character models, with possibly different skeleton topology on top of different skin meshes. Contact being, as recently shown [Basset et al. 2022], a point of utmost attention, the need to preserve them makes motion transfer a really challenging task.

Because of this complexity, some methods restrict the problem to limit the diversity of characters. For instance, in [Basset et al. 2020; Jin et al. 2017], both the source and target characters need to come from the same deformable mesh template, in order to make correspondence easier. The last few years have seen the emergence of learning-based approaches, leveraging deep neural networks to ease generalization compared to optimization-based methods [Aberman et al. 2020; Lim et al. 2019; Villegas et al. 2018]. However, these approaches only take into account the skeleton, neglecting all the information contained in the mesh. [Zhang et al. 2023b] uses the information of the target mesh to minimize self-penetration issues, regardless of contacts appearing on the source mesh. By design, all of those methods will fail to accurately capture important aspects of the motion, and offer a limited degree of generalization to new characters with unseen skeletal structures or morphologies.

Other methods based on reinforcement-learning [Reda et al. 2023; Zhang et al. 2023a] have allowed for a physically accurate retargeting that takes into account mesh contacts, but require training a new agent for each new target character.

Our method represents motion as semantic matrices extracted from the positions of a low number of key vertices on the mesh, generalizing to meshes the work done on skeletons in [Zhang et al. 2023b]. We focus on making them as sparse as possible in both spatial and temporal domains, allowing us to be orders of magnitude faster than previous optimization methods, while retaining a high degree of control and increasing generalization.

2 PROPOSED METHOD

We present a new representation of poses via sparse motion semantics, enabling us to formulate losses that account for interactions such as collision avoidance, surface contacts, and deformations that were previously overlooked by other methods.

MIG '23, November 15–17, 2023, Rennes, France

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

This is done by selecting a subset of N mesh vertices to provide a spatially sparse representation of the morphology of the character, as shown in Figure 1. This step can be manually performed by an untrained human, but optimal transport methods could help make it automatic [Solomon et al. 2016].

We formulate the retargeting task as an optimization problem, where $\mathbf{q}, p$ are the target poses and root bone position :

$$\min_{\mathbf{q},p} \quad w_{\text{sem}}\mathcal{L}_{\text{sem}} + w_{\text{reg}}\mathcal{L}_{\text{reg}} + w_{\text{smooth}}\mathcal{L}_{\text{smooth}} \tag{1}$$

The regularization loss $\mathcal{L}_{\text{reg}}$ penalizes solutions that are far from the initialization, and the smoothness loss $\mathcal{L}_{\text{smooth}}$ ensures that the changes in velocity are small. Unlike previous methods, the semantic loss $\mathcal{L}_{\text{sem}}$ makes use of the distance and direction between pairs of keypoints, in addition to the velocity, normal, and height of each keypoint, to estimate how faithful the retargeted pose is to the source motion. Calling k_i the 3D position of the i^{th} sample point, $\mathcal{L}_{\text{sem}}$ is expressed as the weighted sum of the following losses.

- One loss ensuring that the distance between pairs of keypoints is preserved :

$$\mathcal{L}_{\text{pos}} = \sum_{i,j}^{N} \left(\left\| k_i^{\text{source}} - k_j^{\text{source}} \right\| - \left\| k_i^{\text{target}} - k_j^{\text{target}} \right\| \right)^2$$

- One loss ensuring that the direction between pairs of keypoints is preserved (with S_{cosine} being the cosine similarity):

$$\mathcal{L}_{\text{dir}} = \sum_{i,j}^{N} S_{\text{cosine}} \left(k_i^{\text{source}} - k_j^{\text{source}}, k_i^{\text{target}} - k_j^{\text{target}} \right)^2$$

- One loss using the normals n_i of the keypoints to penalize intersecting limbs :

$$\mathcal{L}_{\text{pen}} = \sum_{i,j}^{N} \max \left(0, n_i^{\text{target}} \cdot \left(k_i^{\text{target}} - k_j^{\text{target}} \right) \right)$$

- One loss penalizing keypoints penetrating the floor :

$$\mathcal{L}_{\text{height}} = \sum_{i,j}^{N} \max \left(0, -k_i^{\text{target}} \cdot \overrightarrow{\text{up}} \right)$$

- A loss penalizing foot sliding across successive frames :

$$\mathcal{L}_{\text{sliding}} = \sum_{t} \sum_{i \in \text{Feet}} \left\| \left(k_{i,t+\Delta t}^{\text{target}} - k_{i,t}^{\text{target}} \right) - \left(k_{i,t+\Delta t}^{\text{source}} - k_{i,t}^{\text{source}} \right) \right\|^2$$

These losses do not apply to all joints for each frame. Instead, we compute a relevance matrix for each of those, based on the source keypoints, in order to ponderate those losses depending on the pose. These matrices are very sparse both spatially (only a few joints are significant for a given frame) and temporally (a given joint is often significant for only a few frames at a time).

3 EXPERIMENTS AND RESULTS

We implemented our computations in Pytorch and used the Adam optimizer to perform gradient descent, processing all frames simultaneously with $N = 42$. Around 30 iterations are needed to yield good results, which gives an inference speed of 25 frames per second on a Nvidia RTX 3060. Visual examples of our retargeting method can be found in Figure 1 and in supplementary material.

We evaluated our method on animations from the Mixamo dataset [Adobe 2023], similarly to [Aberman et al. 2020; Villegas et al. 2018; Zhang et al. 2023b]. We compared our method to state-of-the-art retargeting from [Zhang et al. 2023b], which, by design, does not capture source contacts, nor take ground contacts into account. Additional work is needed to provide quantitative comparison with previous methods, as well as a preferential user study.

Performance-wise, our approach is faster than other optimization-based retargeting work, which reported speed of 0.7 seconds per frame [Jin et al. 2017] and up to 15 minutes per frame [Basset et al. 2020].

4 CONCLUSION AND FUTURE WORK

We presented an efficient contact-aware retargeting method, based on sparse semantics. Even though we focused on humanoid characters, our method could easily adapt to other types of characters, such as quadruped animals. In future work, we will experiment with new losses adapted for more specific tasks, such as multi-character retargeting or interactions with objects. Using another sparse loss for inter-character interactions, our approach would likely scale up to crowds with minimal impact on performance. We would also like to experiment with non-flat or deformable ground surfaces. Finally, learning the sparse weighting matrices could yield even better results.

REFERENCES

Kfir Aberman, Peizhuo Li, Dani Lischinski, Olga Sorkine-Hornung, Daniel Cohen-Or, and Baoquan Chen. 2020. Skeleton-Aware Networks for Deep Motion Retargeting. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 62.

Adobe. 2023. Mixamo. https://www.mixamo.com/. Accessed: 2023-09-12.

Jean Basset, Badr Ouannas, Ludovic Hoyet, Franck Multon, and Stefanie Wuhrer. 2022. Impact of Self-Contacts on Perceived Pose Equivalences. In *Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Guanajuato, Mexico) *(MIG '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 10 pages. https://doi.org/10.1145/3561975.3562946

Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon. 2020. Contact preserving shape transfer: Retargeting motion from one shape to another. *Computers & Graphics* 89 (2020), 11–23. https://doi.org/10.1016/j.cag.2020.04.002

Taeil Jin, Meekyung Kim, and Sung-Hee Lee. 2017. Motion Retargeting to Preserve Spatial Relationship between Skinned Characters. In *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation* (Los Angeles, California) *(SCA '17)*. Association for Computing Machinery, New York, NY, USA, Article 25, 2 pages. https://doi.org/10.1145/3099564.3106647

Jongin Lim, Hyung Jin Chang, and Jin Young Choi. 2019. PMnet: Learning of Disentangled Pose and Movement for Unsupervised Motion Retargeting. In *British Machine Vision Conference (BMVC)*.

Daniele Reda, Jungdam Won, Yuting Ye, Michiel Panne, and Alexander Winkler. 2023. Physics-based Motion Retargeting from Sparse Inputs. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 6 (08 2023), 1–19. https://doi.org/10.1145/3606928

Justin Solomon, Gabriel Peyré, Vladimir G. Kim, and Suvrit Sra. 2016. Entropic Metric Alignment for Correspondence Problems. *ACM Trans. Graph.* 35, 4, Article 72 (jul 2016), 13 pages. https://doi.org/10.1145/2897824.2925903

Ruben Villegas, Jimei Yang, Duygu Ceylan, and Honglak Lee. 2018. Neural Kinematic Networks for Unsupervised Motion Retargetting. arXiv:1804.05653 [cs.CV]

Jiaxu Zhang, Junwu Weng, Di Kang, Fang Zhao, Shaoli Huang, Xuefei Zhe, Linchao Bao, Ying Shan, Jue Wang, and Zhigang Tu. 2023b. Skinned Motion Retargeting With Residual Perception of Motion Semantics & Geometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13864–13872.

Yunbo Zhang, Deepak Gopinath, Yuting Ye, Jessica Hodgins, Greg Turk, and Jungdam Won. 2023a. Simulation and Retargeting of Complex Multi-Character Interactions. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 65, 11 pages. https://doi.org/10.1145/3588432.3591491