CIS 565 Final Project Proposal

Dancing to Music in 3D

Han Yan and Weiyu Du

Background and Motivation

Dancing to music is an instinctive move by humans. However, learning to model the music-to-dance generation process is a challenging problem. 1) Measuring correlation between music and dance movements is hard with many factors to consider, such as beats, style. 2) Dance is multimodal, meaning there could be many different possible generations corresponding to one piece of music.

A music-to-dance mapping can have many applications in assisting content generation in arts and sports. Moreover, studying the problem could yield better understanding of cross-model synthesis.

While most of the existing studies tackle the problem in 2D joint space, we feel that learning only 2D movements not only yields partial understanding of human dynamics, but also hurts the generation in terms of aesthetics and realism. The major challenge for 3D generation is lack of accurate training data. Therefore, we plan to use state-of-the-art video human pose reconstruction models to obtain 3D ground truth dance movements first, then implement a music-to-dance generation model based on existing research [4][5][6].

Schedule

Milestone 1: Obtaining 3D training data using off-the-shelf models to reconstruct 3D joints from dance videos.

Milestone 2: Implement music-to-generation model in 3D joint space with PyTorch based on Lee et al. (Neurips 2019). [4]

Milestone 3: Model training and performance analysis.

Final: Prepare for final presentation and demo.

Schedule

PyTorch [1]: Machine learning library that supports GPU tensor operations

SMPL [2]: Skinned Multi-Person Linear Model

Possible third-party codebase for off-the-shelf pose reconstruction from video [3]: https://github.com/akanazawa/human_dynamics

References

[1] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T.,Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z.,Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.:Pytorch: An imperative style, high-performance deep learning library. In: Wallach,H., Larochelle, H., Beygelzimer, A., d'Alch´e-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)

[2] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia) 34(6), 248:1–248:16 (Oct 2015)

[3] Kanazawa, A., Zhang, J.Y., Felsen, P., Malik, J.: Learning 3d human dynamics from video. In: Computer Vision and Pattern Regognition (CVPR) (2019)

[4] Lee, H.Y., Yang, X., Liu, M.Y., Wang, T.C., Lu, Y.D., Yang, M.H., Kautz, J.: Dancing to music. In: NeurIPS (2019)

[5] Tang, T., Mao, H., Jia, J.: Anidance: Real-time dance motion synthesize to the song. In: 2018 ACM Multimedia Conference on Multimedia Conference. pp. 1237–1239. ACM (2018)

[6] Tang, T., Jia, J., Mao, H.: Dance with melody: An lstm-autoencoder approach to music-oriented dance synthesis. In: 2018 ACM Multimedia Conference on Multi-media Conference. pp. 1598–1606. ACM (2018)