Explaining rehearsals of a symphony orchestra in virtual reality

Michael Zöllner
Institute for Information Systems at Hof University
michael.zoellner@hof-university.de

Markus Bosl
Hof University
markus.bosl@hof-university.de

Dirk Widmann
Hof University
dirk.widmann@hof-university.de

Moritz Krause
Institute for Information Systems at Hof University
moritz.krause.2@iisys.de

Jan Gemeinhardt
Institute for Information Systems at Hof University
jan.gemeinhardt@iisys.de

ABSTRACT

The project Symotiv's (http://symotiv.de) goal is to describe the complex function of a symphony through an interactive experience of the orchestra's digital twin. We are capturing the musicians, their instruments and their sounds with machine learning based 3D human pose estimation and visualize them in Virtual Reality in a three-dimensional, virtual concert hall. Our audience steps into the role of the conductor and experience the scene and the sound from a perspective within the orchestra. This opens the possibility of experiencing and understanding what is happening from new spatial, but above all sonic perspectives.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics • **Networks** → Network reliability

KEYWORDS

ACM proceedings, text tagging

1 INTRODUCTION

Current 3D human pose estimation solutions through machine learning and visualization in virtual reality (VR) enabled us to capture and visualize the complex structure of the Hof Symphony Orchestra. With the help of both technologies, we created an interactive immersive extended reality experience that showed the diverse aspects from rehearsal to performance to a broad audience.

On the one hand, this gives the audience an unfamiliar insight into the cosmos of the orchestra; on the other hand, musicians who themselves play in the orchestra can experience the orchestra from a different perspective. Music students, who rarely have the opportunity to play in an orchestra, can experience the original conditions under which they will later work. Another area of application is the imparting of the orchestra experience to school classes with mobile equipment as well, which goes far beyond what can be taught via textbook.

In the next sections we are describing our approach to capture and to visualize a rehearsal of a symphony orchestra and the resulting music in virtual reality.

2 RELATED WORK

2.1 Machine Learning Based Human Pose Estimation

We observed a fast progress in the quality and availability of camera-based pose estimation over the last years. Machine learning based approaches eliminated the need for special hardware in favor of standard cameras. Carnegie Mellon University's OpenPose [4] and OpenPifpaf [8] are two major developments producing robust 2D pose estimation of humans based on single RGB images. Today's OpenPose version features 3D position data as well in the BODY_25 pose topology.

Like OpenPose there's MeTRAbs Absolute 3D Human Pose Estimator [13] which is featuring not only 2D but also robust 3D position data of the skeletons' joints in the SMPL-24 topology.

Google MediaPipe [10] combines a large variety of machine learning (i.e. pose estimation, hand and gesture recognition, object detection) applications in one framework. Part of the framework is BlazePose [2] which quantifies physical exercises like yoga, dance, and fitness applications and delivers 33 3D landmarks in COCO [9] topology superset GHUM3D and background segmentation mask.

In our prior research we recognized OpenPose performed best regarding accuracy and correct approximation of hidden body parts like musical instruments. Therefore, we were choosing OpenPose in this project for tracking the musicians' motions.

2.3 Related Projects

Regarding related applications we would like to mention several projects in adjoining disciplines. Capturing and analyzing dance is a wide research field in the pose estimation and tracking area. Referring to hardware-based approaches we would like to mention “Evaluating a dancer's performance using Kinect-based skeleton tracking” [1]. An early 3D position estimation approach with a single camera was “Capturing of contemporary dance for preservation and presentation of choreographies in online scores” [7] where they showed recurring path patterns in experimental ballet performances. The former mentioned paper is part of the larger Motion Bank [6] project initiated by dance choreographer William Forsythe. Several universities and research institutions worked on solutions for digital preservation of contemporary dance and ballet. The project's radius has been increased by the Choreographic Coding Lab [11]format that invited artists to work on the topic in workshops in several cities worldwide.

Although focusing on visualization rather than data analysis we would like to mention RunwayML's [12] latest Motion Tracking tool for video editing and Najeeb Tarazi's application of RunwayML's rotoscoping techniques in the impressive “One More Try” [16] experimental skate video.

3 DATA AQUISITION

3.1 Camera Setup

We had to develop a setup to capture 50 musicians and the conductor without occlusion. Therefore, we used eight affordable off the shelf GoPro Hero 5 Black cameras (see figure 1) on long tripods and fixed with a 3D printed mount. One camera for each instrument group.

Figure 1: View of musicians in a instrument group.

The arrangement of the musician's positions and chairs was fixed so that the cameras could capture the group without occlusion of other people, instruments or chairs (see figure 2 and 4).

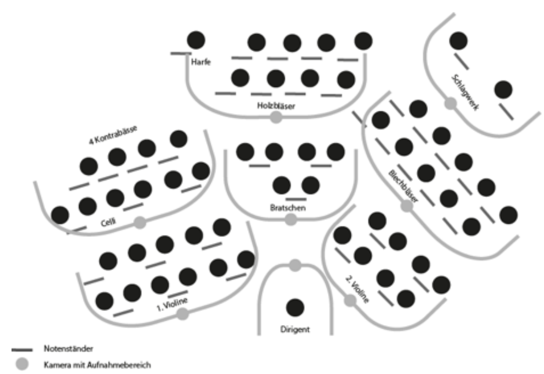

Figure 2: Plan of the musicians in their instrument groups and the position of the cameras.

3.2 Human Pose Estimation Data

After analyzing the videos of the 8 camera tracks OpenPose delivered single JSON with information about the musicians' skeleton joints in the BODY_25 pose topology (see figure 3). A JSON file contains the pose data of all visible humans in the current frame. This resulted in 8 cameras multiplicated by 25 frames per second and the length of the rehearsal.

For the process of rigging 3D characters in Blender [3] and the real-time visualization in Unity3D [14] we wrote a custom script to convert the data into the Biovision Hierarchy (BVH) character animation file format. Since all cameras and musicians were set up to fixed positions, we were able to join the 3D datasets of the different cameras and instrument groups into the real 3D positions of the whole orchestra in the hall.

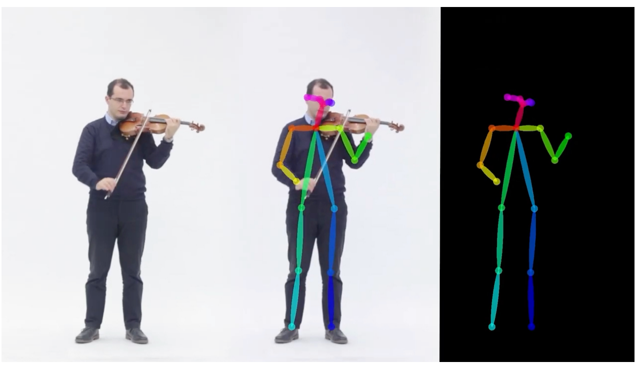

Figure 3: Skeleton data overlayed on the video of a violinist.

4 VISUALIZATION IN VR

We chose the medium virtual reality because of the importance of the venue and the space in which the orchestra plays. For the visualization of the motion data, we designed an abstract room that can be entered and traversed by the users completely freely. Within this space we visualized and explained the collected data and the underlying processes of a symphony orchestra.

4.1 Visualization and Rendering

The interactive 3D framework for our real-time VR visualization was Unity3D running on a standard graphics workstation with two Nvidia GeForce RTX 2080 [5] graphics cards and an HTC Vive Pro VR headset [15] for the VR reexperience of the recordings of the rehearsals.

We converted the motion data into Biovision Hierarchy (BVH) character animation file format to import into Blender with a custom script. Since the cameras and the musicians positions were fixed, we were able to join the 3D datasets of the different cameras and instrument groups into the real 3D positions of the whole orchestra in the hall.

Since we only captured the skeletons of the musicians, we had to design a new visual representation appearance. We chose an abstract retro-futuristic appearance in the style of Disney's Tron [17] movie. The musicians were grouped in their instrument groups like violins, contrabass, brass, etc. and color coded so the users could identify them. Furthermore, we amplified the identification of the instruments group using typography and pictograms of the instruments (see figure 4).

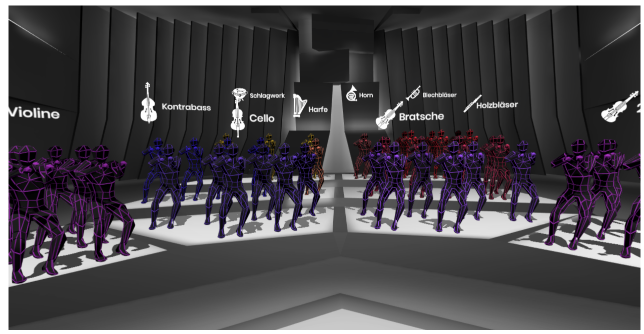

Figure 4: Virtual musicians separated in their instrument groups.

4.2 Interacting with the Orchestra in VR

While the audience in a real concert is sitting and watching from a fixed seat, in our application the users can interactively intervene in the performance. At the beginning of the VR experience there is a short introduction and a tutorial scene about how to move and interact in VR. Afterwards their journey starts in front of the conductor's position in the virtual hall with a view over the whole virtual orchestra. Now they can move freely around the concert hall, stand in-between the musicians and the conductor. Users are able to activate and deactivate instrument groups (see figure 5). Standing among the musicians, they can observe their movement, sound and thus expand and understand the spatial and sonic experience of a concert. The immersive nature of this medium offers scope for new, inclusive information delivery. Various visualizations here reach the user on multiple levels. Like the event of a concert, VR offers an independent form of experience.

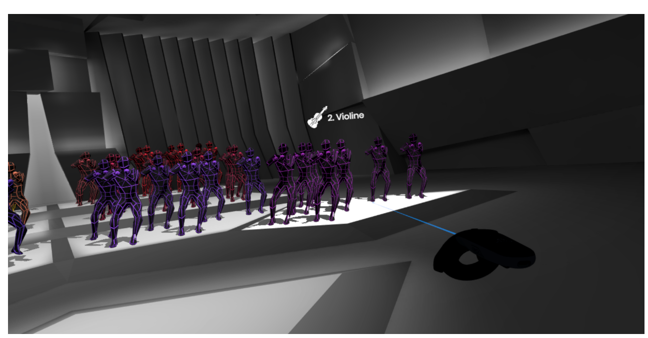

Figure 5: Interacting with the instrument groups.

5 CONCLUSIONS AND FUTURE WORK

In our paper we presented our approach to capture and visualize a symphony orchestra and its music in virtual reality. We presented a solution to describe the complex process of producing classical music to their audience by experiencing the orchestra from normally unreachable positions, observing the motions of human and instrument and directly interacting with the sound tracks. For musicians and the conductor this VR experience is also a tool to improve their skills by replaying the rehearsal in variable positions and focus on different instrument groups.

In our further research we will transfer this concept to different sport disciplines with the aim to explain the interaction of the biomechanical model, forces and figures.

REFERENCES

[1] Alexiadis, D.S., Kelly, P., Daras, P., O'Connor, N.E., Boubekeur, T. and Moussa, M.B. 2011. Evaluating a dancer's performance using kinect-based skeleton tracking. *Proceedings of the 19th ACM international conference on Multimedia* (New York, NY, USA, Nov. 2011), 659–662.

[2] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F. and Grundmann, M. 2020. BlazePose: On-device Real-time Body Pose tracking. arXiv.

[3] Blender Foundation 2023. blender.org - Home of the Blender project - Free and Open 3D Creation Software. *blender.org*.

[4] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. arXiv.

[5] GeForce RTX 20 Series Graphics Cards: 2023. *https://www.nvidia.com/en-us/geforce/20-series/*. Accessed: 2023-02-14.

[6] Hennermann, C. and deLahunta, S. 2013. *Motion Bank: starting points & aspirations = Motion Bank : ansatzpunkte & intentionen*. Motion Bank/The Forsythe Company.

[7] Kahn, S., Keil, J., Müller, B., Bockholt, U. and Fellner, D.W. 2013. Capturing of contemporary dance for preservation and presentation of choreographies in online scores. *2013 Digital Heritage International Congress (DigitalHeritage)* (Oct. 2013), 273–280.

[8] Kreiss, S., Bertoni, L. and Alahi, A. 2021. Openpifpaf: Composite fields for semantic keypoint detection and spatio-temporal association. *IEEE Transactions on Intelligent Transportation Systems*. (2021).

[9] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. 2014. Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (2014), 740–755.

[10] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M.G., Lee, J., and others 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*. (2019).

[11] MOTION BANK: *http://motionbank.org/de.html*. Accessed: 2023-02-14.

[12] Runway - Next-generation creation suite: 2023. *https://runwayml.com*. Accessed: 2023-01-19.

[13] Sárándi, I., Linder, T., Arras, K.O. and Leibe, B. 2021. MeTRAbs: Metric-Scale Truncation-Robust Heatmaps for Absolute 3D Human Pose Estimation. *IEEE Transactions on Biometrics, Behavior, and Identity Science*. 3, 1 (Jan. 2021), 16–30. DOI:https://doi.org/10.1109/TBIOM.2020.3037257.

[14] Unity Real-Time Development Platform | 3D, 2D VR & AR Engine: 2023. *https://unity.com/*. Accessed: 2023-02-12.

[15] VIVE Pro: *https://www.vive.com/us/support/vive-pro/*. Accessed: 2023-02-14.

[16] 2022. *One More Try - an experimental skate video*.

[17] 1982. *Tron*. Walt Disney Productions.