Visualizing Discrepancy of Hand Postures for Performing Piano

Ruofan Liu* Erwin Wu*† Chen-Chieh Liao* 西岡勇人† 古屋晋一† 小池英樹*

概要. Hand postures and finger movements are the most essential factor for learning many dexterous skills, such as playing the piano. However, video teaching is limited by the fixed camera angle, a 2D video is difficult to tell precise 3D hand posture intuitively. This demo presents a real-time visualization system providing the intuitive discrepancy of hand postures between teacher and user. Through a motion capture system, the estimated 3D hand postures are visualized in real time and discrepancies based on distinct metrics are displayed, integrated with modular functions assisting skill acquisition.

1 Introduction and Related Work

Most of the skill acquisitions are based on motion imitation hence correct poses are essential. Especially in some dexterous skills such as playing the piano, understanding a precise finger posture is proved to be beneficial to the learning efficiency [1]. However, current online-piano-learning are conducted through 2D videos, where an intuitive comparison between finger posture is not possible. Liu et al. [2] proposed a training system by visualizing differences between hands in a virtual world, but their system only consists of relatively simple functions. On the other hand, Wang et al [4] introduced the PPVR system where a student can follow an instructor intuitively in virtual reality. However, no further feedback are provided for training.

Based on these previous works, we present a real-time piano self-training system that reconstructs 3D representations and provides visual feedback on hand pose discrepancies. The 3D hand postures are first estimated from the videos and visualized in real time. And a modular UI of different visualizations is provided to the users to observe their finger postures and understand the differences. To validate the proposed approaches, a pilot study is conducted. The results suggest a significant improvement in using the proposed system over conventional video training.

* Tokyo Institute of Technology
† Sony Computer Science Laboratories

2 Implementation

2.1 Apparatus

We used a Logitec Brio web camera and an IR-based depth camera Ultraleap 3Di for video capturing. The visualization system is implemented in Unity3D with a high-end notebook PC. For the user interface, a touch display is used to show visual feedback and to be operated by the user.

2.2 Hand Tracking

As mentioned before, a precise 3D finger position is required for intuitive comparison. We utilize a computer vision-based hand pose estimation model, InterHand [3], which can precisely estimate 2D and 3D dual hand poses from a single webcam using deep neural networks. To enhance the tracking precision of wrist position, the depth camera is also aligned with the RGB camera.

2.3 Visualization

Figure 1 shows the overview of the system 1.

General Functions: Aimed to evaluate the preference for various visualization functions, we integrate them into a modular interface. The timeline and speed controller are similar to that in the ordinary video-teaching method, moving the timeline cursor to decide when to start replay and the speed ranges from 0 to 2 times.

Discrepancy Metrics: To measure the differences between the recorded (student) and tutorial (teacher) data more comprehensively and holistically, at every frame, we have three discrepancy metrics, Corresponding Joint Interval, Corresponding Finger Slope, and Adjacent Bone Angle, respectively. Besides, our system displays the discrepancies between student and teacher's

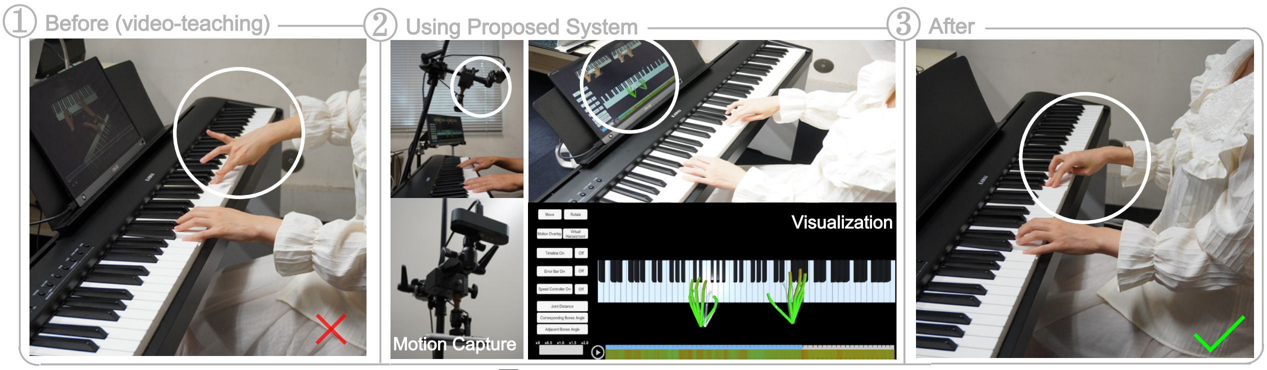

図 1. System Overview

performance in two aspects, i.e. on particular hands visualization and the error bars. In terms of hand visualization, it is clear that all the discrepant areas should be explicit, so the color of different parts (including joints and bones) changes, ranging from green to red. On the other hand, a series of error bars will be shown on the timeline, where an error bar represents one error moment. We use it to emphasize the mistake-made frames in the whole performance and indicate where and what specified kinds of errors are located.

Visualization: Also, we design two different virtual keyboard emplacement, straight-forwardly placing the student's hand and teacher's hand in a motion-overlay manner, and side-by-side placing the keyboards to check and compare the two performances individually.

3 Pilot Study

We recruited ten novices (8 M, 2 F) for a within-subjects user study. They are told to complete a task in two distinct conditions: a baseline conventional video learning and the proposed visualization system. After experiencing each condition for 10 minutes, their performances are recorded for a later quantitative calculation. Finally, a SUS questionnaire is offered and the participants provide their overall comments during the interview.

We calculate the disparity of Mean Per Joint Position Error (MPJPE) between the moment just completing warm-up task and the moment after a teaching condition experience. The visualization system had statistically significantly (t(18)=2.083, p=0.044) lower MPJPE ($-0.69 \pm 0.48mm$) at the end of every 10-minutes piano performance practice compared to video-teaching method ($-0.34 \pm 0.55mm$). SUS was used to perceive usability, and the average score is 68. The score calculated from our questionnaire was 72.75, which was above average and acceptable. From the perspective of the interview conducted, the majority of users gave a high evaluation for error bars providing the discrepancies between their recorded data and tutorial. They note that it can help them quickly locate the mistakes, which reduces their load of recognition and observation to find the errors so that they can directly focus on how to fix the discrepancies and benefit a lot from it.

4 Discussion and Conclusion

Our main finding was that this visualization way results in more correct and effective practicing of finger-refined movements. As professional pianists suggest, this is critical to piano performance, especially when considerable movement control is required to accomplish high levels of precision. Almost none of these are currently possible with only video learning methods. However, we found a limitation of our system that reduces the user ' s rhythmic performance. We assume that this is because users pay more attention to 3D posture instead of temporal information.

Our system implements a visualization system to help users discover hand pose discrepancies between different plays. A pilot study with novices suggests that our system outperforms conventional video-watching approaches. The positive feedback from users has already suggested that the presented approach provides a suitable complementary option to current methods in piano performance learning.

参考文献

[1] S. Furuya, M. Flanders, and J. F. Soechting. Hand kinematics of piano playing. *Journal of neurophysiology*, 106(6):2849–2864, 2011.

[2] R. Liu, E. Wu, C.-C. Liao, H. Nishioka, S. Furuya, and H. Koike. Synchronized Hand Difference Visualization for Piano Learning. In *ACM SIGGRAPH 2022 Posters*, SIGGRAPH '22, New York, NY, USA, 2022. Association for Computing Machinery.

[3] G. Moon, S.-I. Yu, H. Wen, T. Shiratori, and K. M. Lee. Interhand2. 6m: A dataset and baseline for 3d interacting hand pose estimation from a single rgb image. In *European Conference on Computer Vision*, pp. 548–564. Springer, 2020.

[4] X. Wang. PPVR: Teaching and Learning to Play the Piano in Virtual Reality*. UNIVERSITY OF MICHIGAN, 2020.

Future Vision

To better discover where the student's hands differed from the teacher's during the whole performance, it's indispensable to analyze the positions and angles of their fingers at each unit moment. Since the capability between a novice and an expert is quite different for various rhythmical skills like playing the piano, it is technically not possible to compare the corresponding hand pose at every frame. In our future work, we'll align the moments when the two performers are supposed to show the same hand posture. If so, the alignment-based motion difference feedback method for providing the discrepancies for piano learning will definitely helps a lot.

From the pilot study and result above, the positive feedback from participants suggested our method as a kind of "cross-validation" in a real piano-practice situation. It plays an assistant role in learning the piano. Also, we believe that in other skills acquisition, the visualization of pose discrepancy makes sense. Thus this system can also be applied to other motor skills where a correct posture is essential.