Augmented Piano Performance using a Depth Camera

Qi Yang
Computer Science & Engineering Division
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
yangqi@umich.edu

Georg Essl
Electrical Engineering & Computer Science and Music
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
gessl@eecs.umich.edu

ABSTRACT

We augment the piano keyboard with a 3D gesture space using Microsoft Kinect for sensing and top-down projection for visual feedback. This interface provides multi-axial gesture controls to enable continuous adjustments to multiple acoustic parameters such as those on the typical digital synthesizers. We believe that using gesture control is more visceral and aesthetically pleasing, especially during concert performance where the visibility of the performer's action is important. Our system can also be used for other types of gesture interaction as well as for pedagogical applications.

Keywords

NIME, piano, depth camera, musical instrument, gesture, tabletop projection

1. INTRODUCTION

Traditional piano keyboard excels at discrete pitch and note volume controls, and has been part of the interface of most digital synthesizer or sampler instruments. However after the onset of each note, the player have little control of the quality of the sound unless additional controls such as physical knobs or sliders are used. In contrast to bowed or wind-column instruments which have a great range of articulation after sounding each note, piano keyboard provides limited articulations, other than deciding when to stop the note by release the key or the sustain pedal.

In addition, most digital synthesizer or sampler instruments contain more parameters to adjust than acoustic piano, and physical knobs, sliders and switches (or their virtual representations, in case of software instruments) are used to control these parameters. Manipulating these controls during performance can be unintuitive and difficult, especially so if multiple parameters need to be adjusted while notes are sounded.

We used a Kinect depth camera and a video projector to create a 3-dimensional gesture space above the keyboard with the aim to alleviate these shortcomings. By using gesture we hope to make real-time articulatory adjustment easier, as the player can quickly lift their hand from the keyboard and move into and out of the gesture space. With multi-dimensional gestures a single hand can also control multiple parameters rather than trying to manipulate multiple physical controls. Visibility of expressive gestures can be important to convey meaning and nuance to a concert audience. Our system allows good visibility and projection of the articulatory gestures by amplifying the gesture beyond the small range of motion of adjusting knobs or sliders.

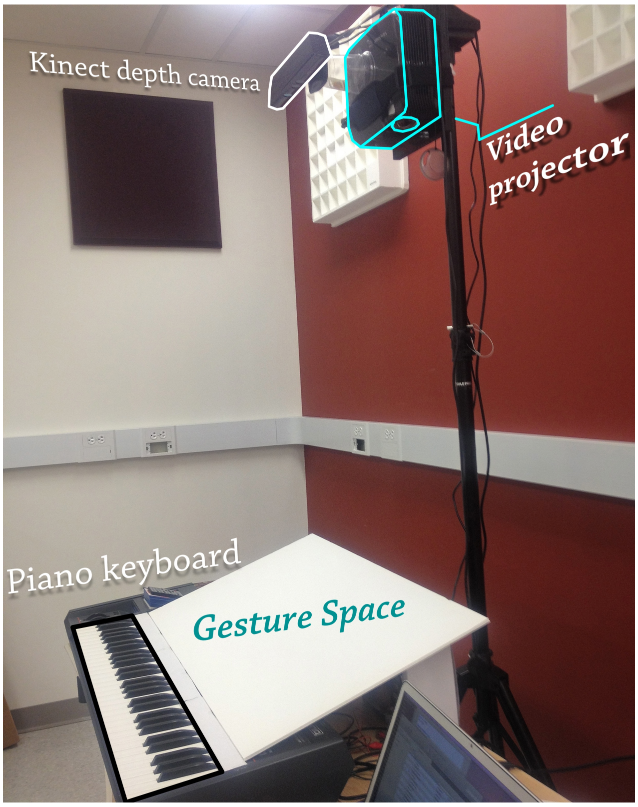

Figure 1: Configuration of the Augmented Piano Keyboard

2. RELATED WORK

Gestures controls are used often for theremin-like music instruments or to augment traditional instruments [4]. Kinect offers an affordable 3D sensing to augmenting acoustic instruments [3]. By retaining the piano keyboard, we hope to retain the advantage of precise and easy-to-learn pitch control, over a purely gesture-based interface.

Works on augmenting the piano keyboard have installed sensors on the keys [2], or on the player's body [1]. Our approach differs by using the depth camera to capture gestures in an open space, without attaching physical sensors to the player or the keyboard, which can be costly or may hinder the normal playing of the keyboard.

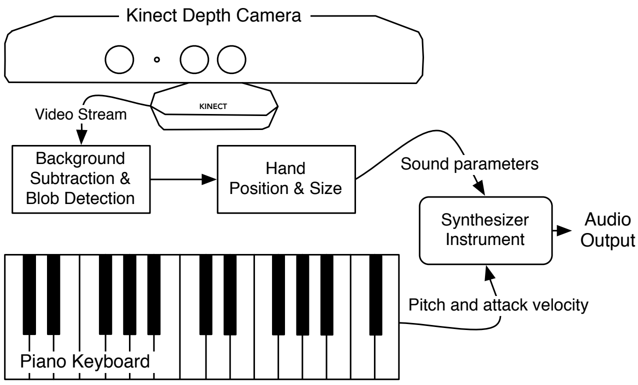

Figure 2: Data flow of the Augmented Keyboard

3. IMPLEMENTATION

The hardware component of our system consists of a Kinect depth camera and a video projector installed above a MIDI piano keyboard, facing down towards the keyboard (Figure 1). The Kinect depth camera, projector and the piano keyboard are all connected to a single computer. The Kinect depth camera captures 3-dimensional data on the gesture space, in the form of a raw video stream. This stream is then passed through background and noise removal and fed into a blob detection algorithm using OpenCV[1]. We can then track the position of the player's hand in 3D, as well as the size of their palm. The hand motion trajectory inferred from this position data is past through an averaging smoothing filter to remove the jitter caused by the noise in the depth camera. Using Processing[2] as a bridge, axes of the hand position data are mapped to different MIDI controller messages that is sent to a software synthesizer (Figure 2).

The Processing framework is also used for displaying visual feedback, which is projected by the projector unto the white surface beneath the gesture space. Figure 3 shows examples of the visual feedback. The detected location of the player's hands are displayed, as well as vertical and horizontal bars signifying the axes that are currently in use and their current values. The circle shows the detected size of the palm as well as the height of player's hand above the gesture surface.

The MIDI piano keyboard is connected to same computer that receives and processes Kinect video stream as well as running the software synthesizer. The synthesizer receives MIDI note pitch and attack velocity messages from the piano keyboard, and articulatory MIDI controller messages such as expression, timbre, or modulation from mapped gesture input.

The piano player can continue to play normally using both hands on the keyboard with audio output provided by the software synthesizer. They can also using one hand to play the notes, while simultaneously lifting the other hand and move it into the gesture space to manipulate the parameters. When gestures are mapped to simple well-understood parameters such as depth of tremolo effect or a high frequency cut-off, we found that the augmented playing style expands the expressive power of piano keyboard.

4. VARIATIONS & APPLICATIONS

Our system is not limited to real-time manipulation of the parameters during playing. The same sensing and visual feedback setup can be adopted for other styles of playing or for applications such as pedagogy.

In one alternative the player uses one hand to press the keys but not sounding the notes, rather activating the pressed pitch classes so that when the gesturing hand passes above the horizontal position of the activated keys, the corresponding note is sounded. This creates the illusion of striking a series of strings in air, akin to the way sets of strings are activated or dampened on harp by pedal, while the harpist runs their finger up the harp, sounding only the activated strings.

In a pedagogical scenario, the hand position data can be used to show contextual information around the learner's hand on the keyboard. For example, a guided improvisation system can show potentially musically sound future harmonies given a history of harmonic progression, by highlighting the appropriate keys to play near the learner's hand. When not used for gesture, the large gesture space can be used to show instructional information, such video or a waterfall representation of the music notation.

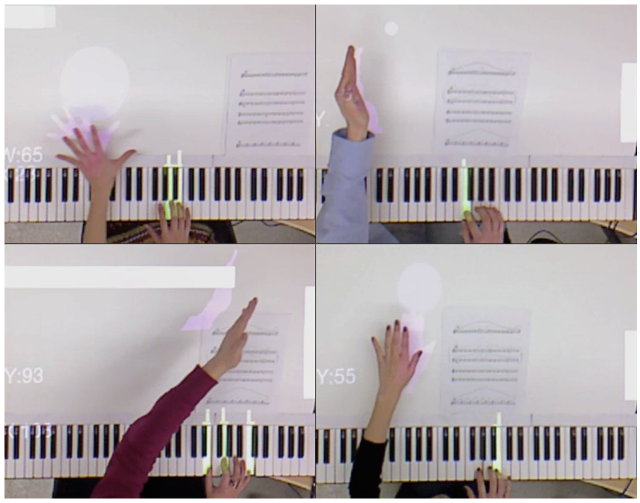

Figure 3: Users illustrating performance gestures

5. FUTURE WORK

We are exploring other performance and pedagogical applications using the same system. We are also conducting a user study on the effectiveness and efficiencies of various gesture-parameter mappings in a performance scenario.

6. REFERENCES

[1] A. Hadjakos and E. Aitenbichler. The elbow piano: Sonification of piano playing movements. In *NIME 2008 Proceedings*, pages 285–288, 2008.

[2] A. McPherson and Y. Kim. Augmenting the acoustic piano with electromagnetic string actuation and continuous key position sensing. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, Sydney, Australia, 2010.

[3] G. Odowichuk, S. Trail, P. Driessen, W. Nie, and W. Page. Sensor Fusion: Towards a Fully Expressive 3D Music Control Interface. In *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, pages 836–841. IEEE, Feb. 2011.

[4] M. M. Wanderley and P. Depalle. Gestural control of sound synthesis. In *Proceedings of the IEEE*, pages 632–644, 2004.

[1]http://ubaa.net/shared/processing/opencv/
[2]http://processing.org/