Open-Source Gesture Recognition for Virtual Orchestra Conducting Using the Kinect

Edward Ly

Earlham College, Department of Computer Science, Richmond, IN

Abstract

- Developing a virtual orchestra conducting application has been a difficult problem in computer music research in part due to the difficulties faced in gesture recognition research.
- Choosing the right hardware for the job is no easy task, and the amount of open-source resources available can wildly vary between each device as well.
- The Microsoft Kinect is one such recent gaming device that has potential to make live virtual conducting possible, but the scarcity of open-source libraries for the Kinect makes development for the device difficult.
- We developed "Kinect Konductor," a virtual conductor application for the Kinect for Windows v1 that is dependent on nothing but open-source, cross-platform libraries.
- With our application, we were able to detect beats with 97.8% accuracy and play music with minimal latency.

Introduction

- We believe the Kinect for Windows v1, despite being released in 2012 and currently discontinued from further production, still paves the way forward for further development into the problem of gesture recognition and virtual conducting.
- While the Kinect for Windows v2, released in 2014, is a stark improvement over the previous version, its incompatibility with existing open-source Kinect libraries proves to be an unfortunate setback for open-source Kinect development.
- Furthermore, the OpenNI framework and NITE middleware were popular resources for Kinect development. However, NITE is proprietary and is no longer available for download.
- Previous research into musical applications using the Kinect have only gone so far as either to just detect beats [3] or to make a custom virtual instrument out of the Kinect [1].
- Our goal was to take all of the available resources we could find and develop a virtual conductor for the Kinect with these resources.
- Along with the use of the MIDI protocol, we ultimately developed an application that could conduct not only orchestral music, but also music of any genre.

Program Design & Implementation

- Key libraries and functions (Fig. 1):
  - libfreenect: extract Kinect depth images
  - XKin: isolate hand and track its position [2]
  - OpenCV: digital imaging, GUI (Fig. 2)
  - FluidSynth: send MIDI note messages to internal sequencer and synthesizer
  - JACK: route FluidSynth audio output to system audio device
- Hand Detection & Position Tracking:
  - Isolate body from the rest of the image
  - Isolate hand from body
  - Get hand position point by taking the centroid of the hand contour
- Beat Detection:
  - Watch for change in velocity in the y-direction
  - Use magnitude of change to estimate volume

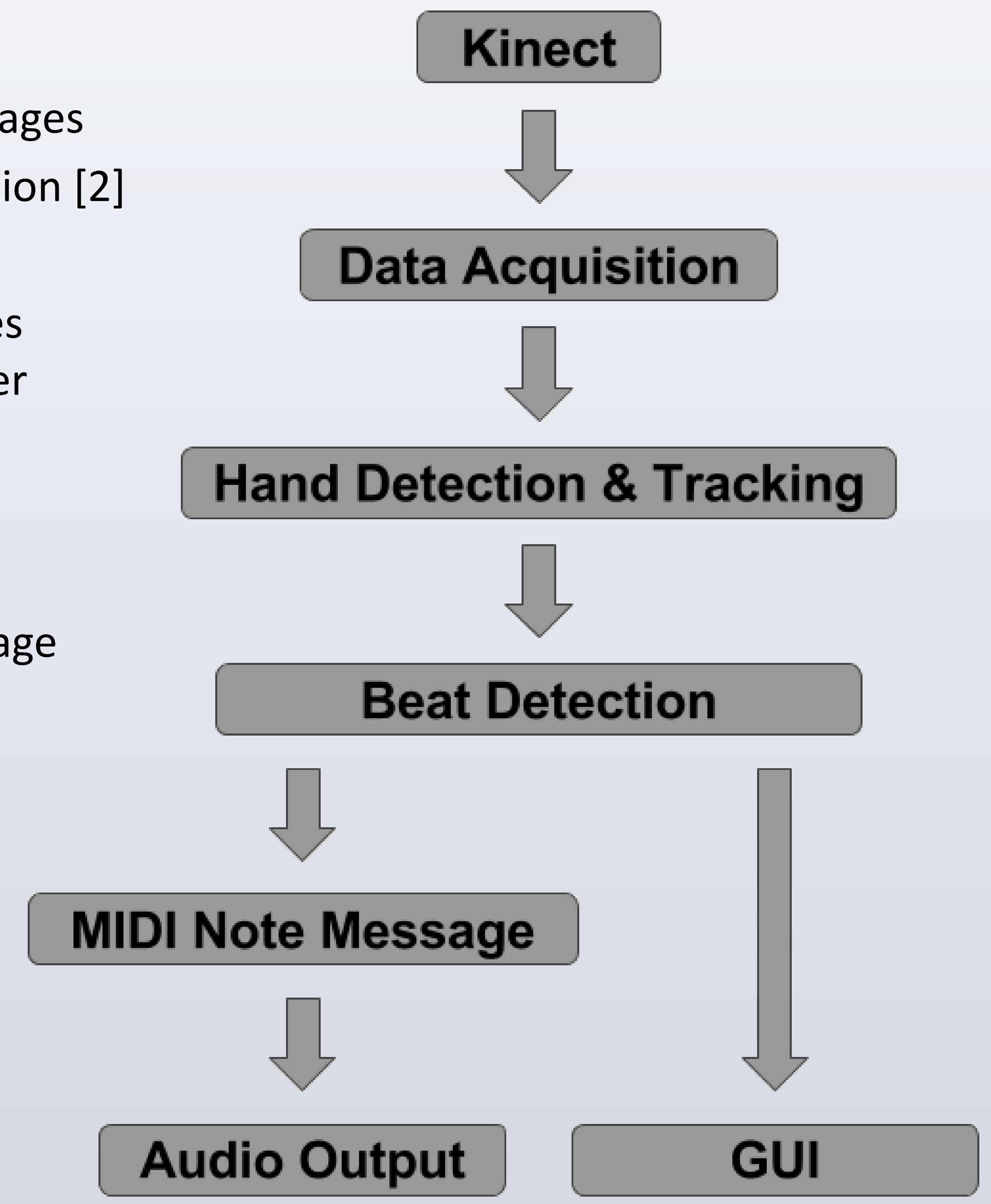

Figure 1: Overview of the flow of data in our program.

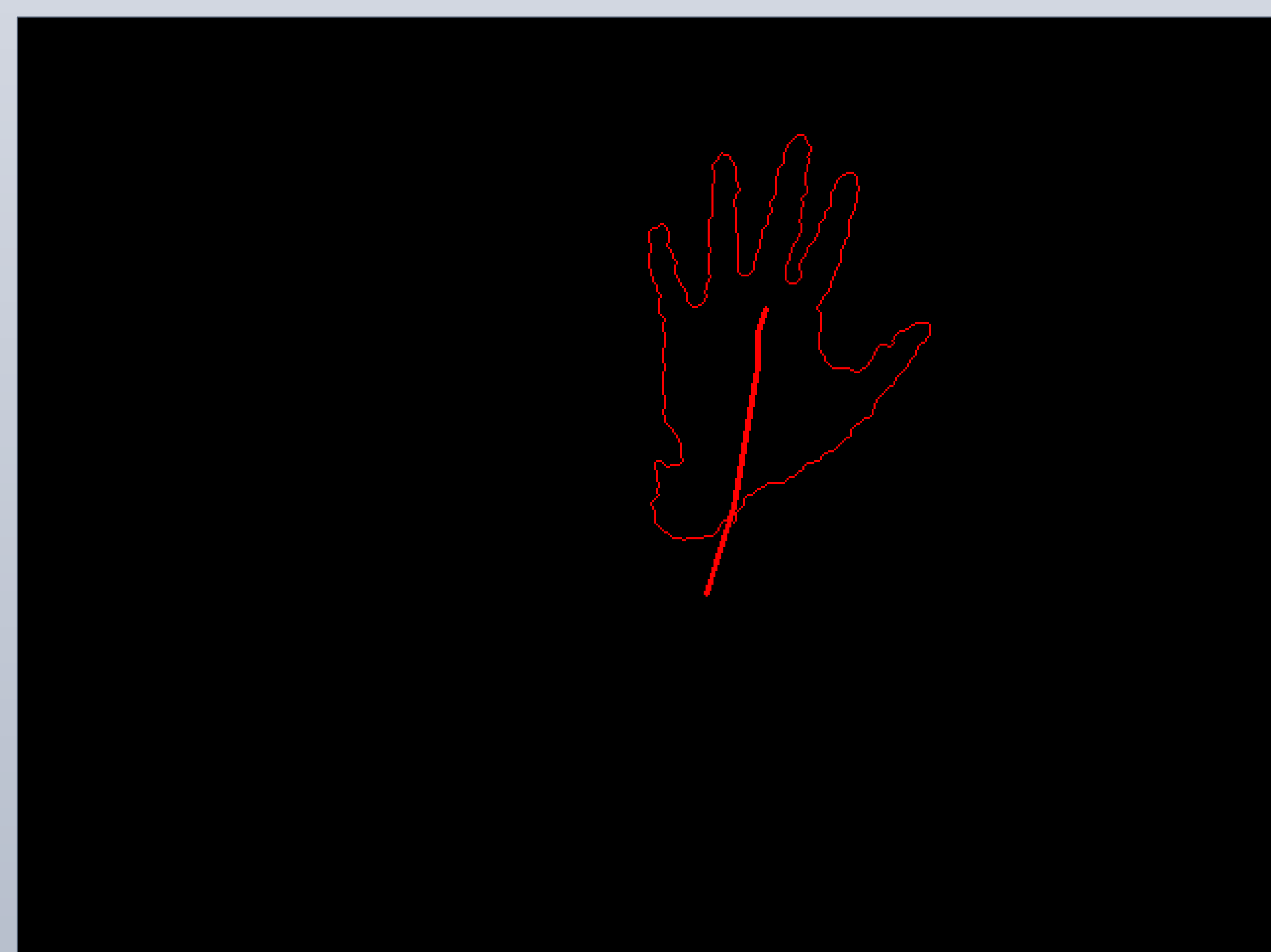

Figure 2: A screenshot of the program GUI.

Results

- Two key performance metrics: beat detection accuracy and latency.
- Latency: time from beat gesture performed to corresponding audio output.
- Average accuracy of our program was 97.8%, or about one mistake every 48 beats on average.
- Latency was too small to notice a significant delay.

Future Work

- Improve algorithms for hand detection, gesture classification, and beat detection.
- Add the ability to vary the volume of each instrument individually.
- Improve the GUI by adding text and other information for users.
- Add cross-compatibility for Windows and Mac computers.

Acknowledgements

- Forrest Tobey and Charlie Peck for their advising and support for this project.
- Xunfei Jiang and Sierra Newby-Smith for their feedback and help during the research and writing of this project.
- Fabrizio Pedersoli for developing the XKin library and providing the necessary tools to make our application a reality.

References

[1] K. Graham-Knight and G. Tzanetakis. Adaptive Music Technology Using the Kinect. In Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '15, pages 32:1–32:4, New York, NY, USA, 2015. ACM.

[2] F. Pedersoli, N. Adami, S. Benini, and R. Leonardi. XKin - Extendable Hand Pose and Gesture Recognition Library for Kinect. In Proceedings of the 20th ACM International Conference on Multimedia, MM '12, pages 1465–1468, New York, NY, USA, 2012. ACM.

[3] A. Sarasúa and E. Guaus. Beat Tracking from Conducting Gestural Data: a Multi-Subject Study. In Proceedings of the 2014 International Workshop on Movement and Computing, MOCO '14, pages 118:118–118:123, New York, NY, USA, 2014. ACM.