Piano Music Transcription Based on Computer Vision

Robert McCaffrey, Master of Computer Science

University of Dublin, Trinity College, 2017

Supervisor: Dr. Kenneth Dawson-Howe

Current research in the area of automated transcription technology focuses primarily on audio signal processing. The precision of audio signal processing is significantly below that of a human expert due to the presence of polyphonic tones (several sounds occurring simultaneously), and the difficulties of successfully parsing such signals given current algorithms. The aim of this paper is to investigate the application of computer vision to piano transcription. Transcription of music is here defined as creating a piece of music notation, in this case by analysing a video feed, so as to extract the musical parameters of the sounds that occur within the performance.[1] The video feed to be analysed shall be captured through a stationary mounted device that is angled at the side of the piano, while a pianist performs on the piano below. The transcription shall be accomplished purely through computer vision techniques, which shall examine key changes at each piano key in real-time. The results of the computer vision application create an industry standard digital representation of an instrument's performance known as a MIDI (Musical Instrument Digital Interface) file. The MIDI file can then be rendered to written music notation known as sheet music, through freely available external software. The performance of the system worked best on an acoustic piano, and when the rhythm of the piece was fast. The average precision and recall were 78.72% and 93.57% respectively, with two outliers dramatically affecting precision.