Visual Detection of Guitar Fretboard and Subsequent AR Finger Position Overlay Based on Tabulature Data

Hyunbin Kim
Dept. of Business Administration
Seoul Nat'l University
hbkim96@snu.ac.kr

Jinsol Park
Dept. of Liberal Studies
Seoul Nat'l University
jinsolpark@snu.ac.kr

Sue Hyun Park
Dept. of Business Administration
Seoul Nat'l University
brokenopen10@snu.ac.kr

Abstract

The main objective of our project is to utilize the front-facing camera of smartphones to create a computer vision-based AR tool for first-time guitar learners who have a hard time matching tabulature with specific positions on the fretboard. We will attempt to either directly use or simulate the use of: image pre-processing through CV methodologies; basic ML; and various smartphone sensors/functionalities such as LiDAR, ToF, and multiple FoV cameras - in order to make possible real-time, reliable analysis of the fretboard and subsequent projection of finger position as an overlay onto the video. This project will work as a proof-of-concept that such functionalities can be incorporated into a smartphone app for first-time guitarists to utilize as an improved medium of self-education.

1. Introduction

While the guitar is a very popular instrument, properly learning how to play it is a difficult, and in most cases, quite unpleasant task. This is in part due to how guitar scores are designed. The two most popular forms of guitar scores - tabulature and chord progression - are both intrinsically flawed. While tabulature is intuitively understandable and playable, it removes chord, and therefore hand/finger position data, making it hard for first-time guitarists to quickly ascertain exactly how they must hold the frets corresponding to the presented tabs. On the other hand, while chord progression scores preserve chord data, it neglects exact fret positions, requiring extensive knowledge of chords and their variants before becoming playable - which is, again, not considerate of first-time learners.

This is why most first-time guitarists resort to Youtube videos when learning songs. Neither tabulature nor chord progression is usable until the player has basic to intermediate knowledge of playing the guitar. However, video lessons are not ideal when learning an instrument, since videos do not wait for the learner; when learning guitar through Youtube videos, a substantial amount of time is spent pausing and rewinding the video.

These flaws demonstrate the need for an alternate medium of education. Some considerations based on what we've discussed above: for efficient and effective learning, this medium should be visual in nature, so that learners can *see* exactly where and how they must press the fretboard; these visual cues must be given in real-time, as the learner progresses through a song, *i.e.*, it must wait for the learner, unlike a recorded video lesson; and lastly, it must put emphasis on accessibility - it should be accessible at any time, anywhere, and preferably should not require extensive setup or modifications to the instrument in order to work. Based on these considerations, it would be reasonable to assume that smartphones, considering its portability, various camera/vision sensors, and high computing/ML performance capable of supporting vision-related tasks, should act as our preferred medium.

Therefore, in this project, we will attempt to establish a proof-of-concept of a smartphone app capable of real-time visual detection of a guitar fretboard captured with its front-facing camera; which we will then use to project AR-based feedback on the corresponding area of the input video in real time. Possible extensions to functionality include detection of currently pressed frets; overlaying hand position/orientation in addition to fret positions; and analysis of the right hand. We plan to make use of existing methods explained below, and modify them for better applicability to our use case if possible.

2. Relevant papers

2.1. Real-time Guitar Fretboard and Fingering Detection

Burns and Wanderley [1] presented a classic method for visually detecting and recognizing fingering gestures of the left hand of a guitarist in real-time. The input data they choose is produced by a camera mount on the guitar neck, which is not the desired input for our project. Still, their idea about finger movement segmentation, which is to compute the local minima in the motion curve of the guitarist, is notable.

Scarr and Green [4] proposed an improved method that does not require a mounted camera but provides detection with higher accuracy. First, on the input data showing a global view of the fretboard, background subtraction and line detection is applied, and a cluster of lines with specific gradients is assumed to constitute the fretboard. A bounding rectangle is constructed to contain the cluster. Next, to obtain a close-cut view of individual frets, the image bounded by the rectangle goes through line detection, where each cluster of approximately vertical lines is considered to be a fret. The proposed fitting algorithm detects each fret while a scanline algorithm finds the nut that helps estimate fret numbers. Further, the fretboard image is normalized so that the detected frets correspond to their expected locations in "fretspace". Finally, after edge detection on the normalized fretboard image, individual contours are filtered by the proposed criteria to identify the fingers. The contour information is then used to heuristically determine which frets are active in any given frame.

The output of Scarr and Green's prototype was evaluated by manual inspection on the correctness of each detected note. While individual finger positions were detected with a significant increase in accuracy compared to Burns and Wanderley's results, the recognition rate was the worst for chord progression. Five types of failure were spotted (from high frequency): incorrect neck bounds, camera angle error, joined fingers, incorrect linear transform, and finger not detected.

In order to input live video feed from the smartphone's front-facing camera, we plan to comprehensively follow Scarr and Green's pipeline. To extend our task to processing chord progressions that involve condensed hand shapes and transitions in fingering, we can consider improving the temporal segmentation method suggested by Burns and Wanderley. To add, the error types Scarr and Green discover give insight on how to enhance the accuracy of our detection algorithm.

2.2. Guitar Tracking

Motokawa and Saito [3] introduced a guitar tracking methodology for their application of support guitar playing system. Motokawa and Saito initially attempted to attach a marker on the guitar, so that the system could track that marker to locate the guitar. However, this did not work as expected, so edge-based tracking was used.

They additionally sampled points along the projection of the edge on a 3D model, and checked the gradient of the image intensity for each sampled point. This was to locate a 3D augmented reality image of a fixture. However, our project does not need applications to this extent, so we will just apply the edge detection for the guitar.

2.3. Real-time Guitar Chord Estimation

Kerdvibulvech and Saito [2] proposed a method to estimate each guitar chord in real time. They used markers attached to each fingertip of the left hand, and first constructed a database depending on the position of each finger. In other words, for the four fingers (fore finger, middle finger, ring finger, and little finger), they gathered the minimum and maximum x, y, and z position of the fingertip for 100 frames of each guitar chord.

Thus, once the player has played the guitar, the system analyzes the input tracking results (i.e., the input 3D positions of the four finger markers) in each frame. Then, using this input data and the pre-constructed chord-finger database, they estimate the guitar chord as a real-time process. In this process, there is no need to check each feature or run a neural net model to check the position of each fingertip.

If we were to apply this in our project, we will only need the x, and y min, max database since we will be considering a guitar-playing video as a 2D. Using this suggested algorithm will enable real time chord detection, allowing our algorithm to check if the guitarist is playing the correct chord or not.

3. Potential Datasets

Datasets can mostly be generated through our own smartphones/guitars, and edge cases can be obtained through the huge database of guitar lessons and covers available on Youtube. If we decide to utilize other sensors present in a smartphone, we will obtain data, again, using our own smartphones, or simulate the use of them through direct measurements of what the sensor evaluates; *e.g.* distance measurements for LiDAR/ToF sensors. Additionally, to emulate real-world scenarios, we will utilize preexisting online databases of guitar tabulatures (*e.g.* https://www.songsterr.com) and videos of hand positions to try to map actual songs' tabs/chord progressions and their respective hand positions to our input video.

References

[1] Anne-Marie Burns and Marcelo M. Wanderley. Visual methods for the retrieval of guitarist fingering. *Proceedings of the 2006 conference on New Interfaces for Musical Expression*, pages 196–199, 2006. 2

[2] Chutisant Kerdvibulvech and Hideo Saito. Real-time guitar chord estimation by stereo cameras for supporting guitarists. *Proceedings of the International Workshop on Advanced Image Technology*, pages 256–261, 2007. 2

[3] Yoichi Motokawa and Hideo Saito. Support system for guitar playing using augmented reality display. *IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 243–244, 2006. 2

[4] Joseph Scarr and Richard Green. Retrieval of guitarist fingering information using computer vision. *25th International Conference of Image and Vision Computing New Zealand*, pages 1–7, 2010. 2