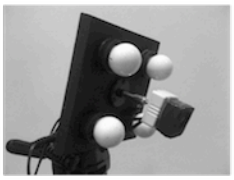
Visual Pitch Class Profile

A Video-Based Method for Real-Time Guitar Chord Identification

M. Cicconet, P. Carvalho, L. Velho (Visgraf / IMPA) & M. Gattass (Tecgraf / Puc-Rio)

We propose a video-based method for real-time guitar chord identification which is analogous to the state-of-the-art audio-based method. While the method based on audio data uses the Pitch Class Profile feature and supervised Machine Learning techniques to "teach" the machine about the chord "shape", we use as feature the approximated positions of fingertips in the guitar fretboard (what we call Visual Pitch Class Profile), captured using especial hardware. We show that visual- and audio-based methods have similar classification performance, but the former outperforms the latter with respect to the immunity to noise caused by strumming.

Capture Hardware



Infrared camera surrounded by four infrared light sources.

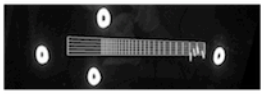
Hollow disk made with retro-reflexive material, four of which are used to locate the ROI.

Middle-phalanges gloves with small rods coated so as to easily reflect light.

Feature Exctraction Pipeline

A threshold is applied to take only guitar and finger markers.

Guitar fiducials and finger rods are detected using a contour detection algorithm.



A projective transformation "immobilizes" the guitar, despite musician movements.

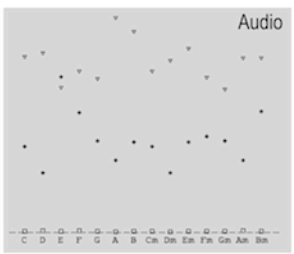
The north-most extreme of finger rods are expressed in guitar-fretboard coordinates.

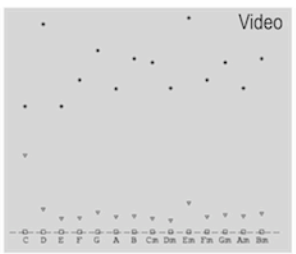
The chord a musician plays is viewed by the system as an eight-dimensional vector composed by the coordinates (after projective transformation) of the four fingertips, from the little to the index finger. We call this eight-dimensional vector the Visual Pitch Class Profile (VPCP).

Chord Identification

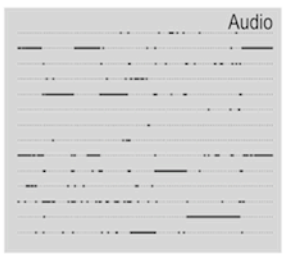
We use a Supervised Machine Learning algorithm: in the training phase, the musician chooses the chords that must be identified and takes some samples from each one of them, where by sample we mean the Visual Pitch Class Profile. In the identification phase, the system receives the vector corresponding to the chord to be identified and classifies it using the chosen algorithm.

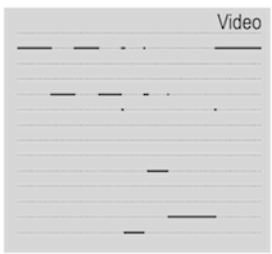
Results





Analysis of clusters. A square (respectively, a triangle) represent the average (respectively, the maximum) distance between the class samples and the class mean vector. The asterisk represent the distance between the cluster mean vector and the nearest cluster mean vector. Video samples provide clusters better defined than audio samples.





The same chord sequence, played twice, is analyzed by the traditional audio-based algorithm and our proposed video-based method. While the former needs some extra processing to cope with the noise caused by strumming, our video based method is immune to that. However, both techniques have problems with chord transitions.