AUDIO-VISUAL GUITAR TRANSCRIPTION

Marco Paleari, Benoit Huet[*]

Multimedia Department
Eurecom Institute
Sophia Antipolis, France

Antony Schutz, Dirk Slock

Mobile Communication Department
Eurecom Institute
Sophia Antipolis, France

ABSTRACT

Music transcription refers to extraction of a human readable and interpretable description from a recording of a music performance. Automatic music transcription remains, nowadays, a challenging research problem when dealing with polyphonic sounds or when removing certain constraints. Some instruments like guitars and violins add ambiguity to the problem as the same note can be played at different positions. When dealing with guitar music tablature are, often, preferred to the usual music score, as they present information in a more accessible way. Here, we address this issue with a system which uses the visual modality to support traditional audio transcription techniques. The system is composed of four modules which have been implemented and evaluated: a system which tracks the position of the fretboard on a video stream, a system which automatically detects the position of the guitar on the first fret to initialize the first system, a system which detects the position of the hand on the guitar, and finally a system which fuses the visual and audio information to extract a tablature. Results show that this kind of multimodal approach can easily disambiguate 89% of notes in a deterministic way.

1. INTRODUCTION

Written music is traditionally presented as a score, a musical notation which includes attack times, duration and pitches of the notes that constitute the song. When dealing with the guitar this task is usually more complex. In fact, the only pitch of the note is not always enough to represent the movements and the positions that the performer has to execute to play a piece. A guitar can indeed chime the same note at different positions of the fretboard on different strings (See Fig. 1). This is why the musical transcription of a guitar usually takes form of a tablature. A tablature is a musical notation which includes six lines (one for each guitar string) and numbers representing the position at which the string.

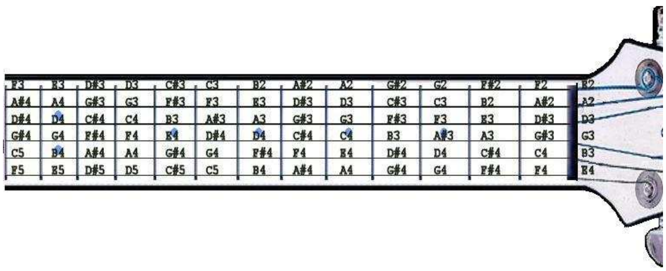

Fig. 1. Notes on a guitar fretboard

Burns and Wanderley [1] report few attempts that have been done to automatically extrapolate fingering information through computer algorithms: real time processing using midi guitar, post processing using sound analysis, post processing using score analysis. Verner [2] retrieves fingering information through the use of midi guitar. Using a midi guitar with different midi channels associated to each different string. Traube [3] suggests a solution based on the timbre. If two notes have the same pitch they can have different timbre. Common issues are precision, needs for a-priori knowledge, and monophonic operation limitation. Another possibility is to analyze the produced score and to extract the tablature by applying a set of rules based on physical constraints of the instrument, biomechanical limitations, and others philological analysis. This kind of methods can result [4] in tablatures which are similar to the one generated by humans, but hardly deal with situations in which the artistic intention or skill limitations are more important than the biomechanical movement. Last but not least, Burns and Wanderley [1] propose to use the visual modality to extract the fingering information. Their approach makes use of a camera mounted on the head of the guitar and extracts fingering information on the first 5 frets but is not applicable to all cases because it needs ad hoc equipment, configuration, and it only returns information about the first 5 frets. This paper presents a multimodal approach to address this issue. The proposed approach combines information from video (webcam quality) and audio analysis in order to resolve ambiguous situations.

[*]Eurecom Institute's research is partially supported by its industrial members: BMW, Bouygues Télécom, Systems, France Télécom, Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The research reported herein was also partially supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space.

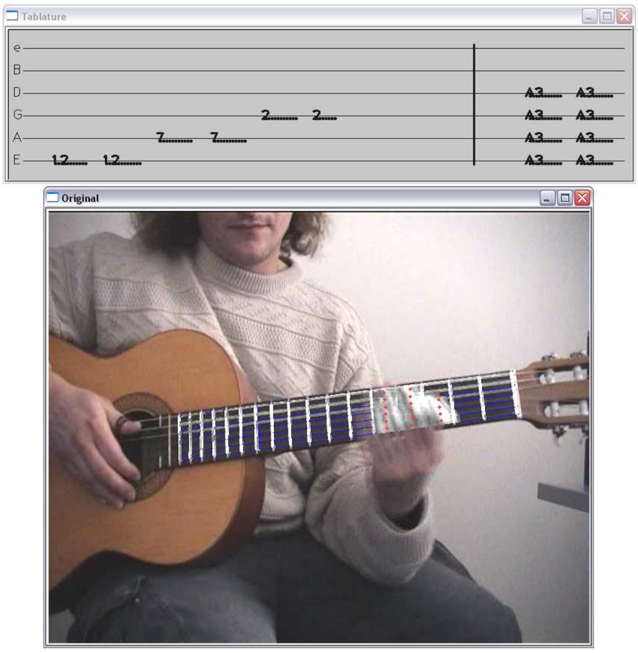

Fig. 2. Interface of the Automatic Transcription System

2. GUITAR TRANSCRIPTION

The typical scenario involved in the discussion of this paper involves one guitarist playing a guitar in front of a web-cam (XviD 640x480 pixels at 25 fps). In the work presented here the entire fretboard of the guitar needs to be completely visible on the video.

2.1. Automatic Fretboard Detection

The first frame of the video is analyzed to detect the guitar and its position. The current version of our system presents few constraints: the guitarist is considered to play a right handed guitar (i.e. the guitar face on the right side) and to trace an angle with the horizontal which does not exceed 90°. The background is assumed to be less textured than the guitar. As a final result, this module returns the coordinates of the corner points defining the position of the guitar fretboard on the video (two outermost points for each detected fret).

2.2. Fretboard Tracking

We have described how the fretboard position is detected on the first frame of the video. We make use of the Tomasi Lukas Kanade algorithm to follow the points along the video.

The coordinates of the end points of each fret are influenced by the movement of the hand. Therefore, some template matching techniques are applied to enforce points to stick to the fretboard. Two constraints were chosen to be invariant to scale, translation or 3D rotations of the guitar: 1) all the points defining the upper (as well as lower) bound of the fretboard must be aligned; 2) the lengths of the frets must comply to the rule $L_i = L_{(i-1)} * 2^{-1/12}$ where L_i represent the length of the i^{th} fret.

To enforce the first constraint a first line is computed that matches the highest possible number of points. The points apart from the line are filtered out and a linear regression (least squares) is computed. All points apart from this second line are filtered out and recomputed.

The second constraint is applied by comparing the positions of the points with a template representing the distances of all the frets from the nut (i.e. the fret at the head of the guitar). Every twenty seconds the tracking is re initialized to solve any kind of issues which may arise from a wrongful adrifts of the Lukas Kanade point tracking.

2.3. Hand Detection

In section 2.2 the methodology employed to follow the position of the frets along the video has been described. Thanks to these coordinates it is possible to separate the region belonging to the fretboard into $n\_strings \times n\_frets$ cells corresponding to each string/fret intersection.

Filtering is done on the frame to detect the skin color and the number of “hand” pixels is counted. A threshold can be applied to detect the presence of the hand (see figure 2).

3. CONCLUSIONS

In this paper we have overviewed a complete, quasi unconstrained, guitar tablature transcription system which uses low cost video cameras to solve string ambiguities in guitar pieces. A prototype was developed as a proof of concept demonstrating the feasibility of the system with today technologies. Results of our studies are positive and encourage further studies on many aspects of guitar playing.

4. REFERENCES

[1] A. Burns and M. M. Wanderley, “Visual methods for the retrieval of guitarist fingering,” in *NIME ’06: Proceedings of the 2006 conference on New interfaces for musical expression*, Paris, France, 2006, pp. 196–199.

[2] J. A. Verner, “Midi Guitar Synthesis: Yesterday, Today and Tomorrow,” *Recording Magazine*, vol. 8 (9), pp. 52–57, 1995.

[3] C. Traube, *A Interdisciplinary Study of the Timbre of th Classical Guitar*, Ph.D. thesis, McGill University, 2004.

[4] D. P. Radicioni, L. Anselma, and V. Lombardo, “A Segmentation-Based Prototype to Compute String Instruments Fingering,” in *CIM04: Proceedings of the 1st Conference on Interdisciplinary Musicology*, 2004.