IMPROVED AUTOMATIC CHORD RECOGNITION

Maksim Khadkevich
FBK-irst, Universitá degli studi di Trento,
Via Sommarive, 14 - Povo - 38050, Trento, Italy
khadkevich@fbk.eu

Maurizio Omologo
Fondazione Bruno Kessler-irst
Via Sommarive, 18 - Povo - 38050 Trento, Italy
omologo@fbk.eu

ABSTRACT

This paper describes a chord recognition system submitted to the MIREX 2009 Audio Chord detection contest. Extracting harmonic information from audio signals has become a topic of keen interest for many researches in Music Information Retrieval (MIR) community. The FBK submission consists of two chord detection system: baseline and the system with language modeling functionality. The two submissions are bases on hidden Markov models (HMM) as a statistical classifier. Pitch class profile (PCP) vectors that represent harmonic information are extracted from the given audio signal and act as a feature set. After Viterbi decoding and subsequent lattice rescoring the output labels are produced.

1. INTRODUCTION

Extracting harmonic structure from raw audio in the form of chord sequence is the task that has been emerging during the past decade. A number of research groups put their attention to the problem of effective and accurate chord recognition. Chord sequence is a high-level feature that reflects harmonic properties of the analyzed signal. It has been successfully used for audio emotion classification [1], cover song identification [2], music structure segmentation [3], audio key estimation [4].

Up to now the most successful and commonly used feature set is considered to be chroma vector (also called pitch class profile) that was introduced in 1999 [5] by Fujishima. Similar to spectrogram that represents spectral content of the signal over time, chromagram (a sequence of chroma vectors) describes the pitch class content. There were some attempts to use features derived from chromagram like tonal centroid [6] and FFT of the chroma vectors [7]. These features are shown to outperform standard chromagram as reported in [6] and [7].

On the stage of feature extraction for a lot of attention has been paid to tuning issues [8–10]. The necessity of tuning appears when audio was recorded from instruments that were not properly tuned in terms of semitone scale. They can be well-tuned relatively to each other, but the reference frequency of ”A4” note can differ from conventional 440 Hz. This mistuning can lead to worse feature extraction and as a result less efficient or incorrect classification. Harte and Sandler [8] suggested using 36 dimensional chroma vectors. In this case they choose the best among 3 possible candidates. Ueda et al. [7] utilized similar approach.

For the moment the best impartial assessment of the chord recognition systems performance is considered to be MIREX [11] competition that is held every year in the context of the ISMIR conference.

The majority of the current state-of-the-art techniques are based on statistical approaches (hidden Markov models (HMM) [12], [13], [7] and dynamic Bayesian network [14]), template matching approaches [15]. Submissions based on the above cited approaches were among the top-ranked results in the MIREX 2009 competition.

In the HMM-based approaches PCP acts as an observation vector. In [13] was shown how N-grams could be useful to model chord sequences and how chord duration information can be incorporated into language modeling. In [14] a 6-layered dynamic Bayesian network was suggested. In this network four hidden source layers jointly model key, metric position, bass pitch class and chord. The two observed layers model bass and treble content of the signal. This paper shows an example of how simultaneous estimation of beats, bass and key can contribute to the chord recognition rate. A fast and efficient template-based chord recognition method was suggested in [15]. The chord is determined by minimizing a measure of fit between the chromagram frame and the chord templates. This system proved the fact that template-based approaches can be as effective as probabilistic frameworks.

The paper is structured as follows: section 2 introduces usage of factored language models for the chord recognition task, functional blocks of the submitted system are presented in section 3, results and discussion are then given in section 4.

2. LANGUAGE MODELING

Modeling chord sequence allows one to improve chord recognition accuracy to some extent (Cheng et al. [1]). In the submitted system the problem of chord progression modeling is addressed through factored language models (FLM) that has been recently suggested for modeling hu-

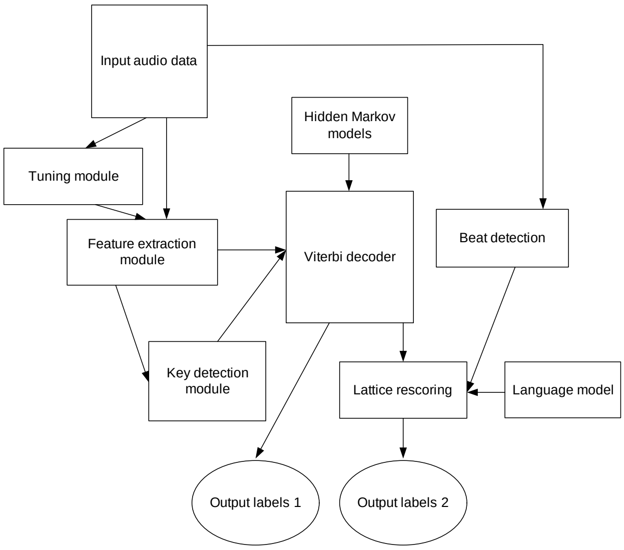

Figure 1. Chord recognition system.

man languages [16, 17] and adapted to the chord recognition task [13]. A factored language model is an extension of N-gram model, where a single unit (chord) can be represented as a bundle of factors. Along with the chord label, a significant amount of information is concentrated in the chord durations. Chord instance can be represented in the following notation: $W-LABEL : D-DURATION$, where the duration length is measured in beats. For chord duration estimation a beat extraction module is used [18]. Language model parameter estimation is performed through training on a database of label transcriptions.

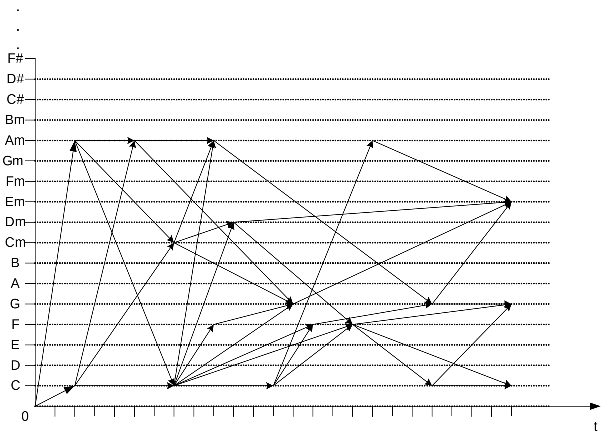

Figure 2. Lattice representation.

3. OVERVIEW OF THE CHORD RECOGNITION SYSTEM

The overview of the whole system designed is depicted in Figure 1. Processing stage of the input audio signal starts with estimating mistuning rate in the tuning block. Algorithm proposed in [19] is adopted here. Ubiquitous chromagram acts as a feature set. Feature extraction process is described in more details in [13].

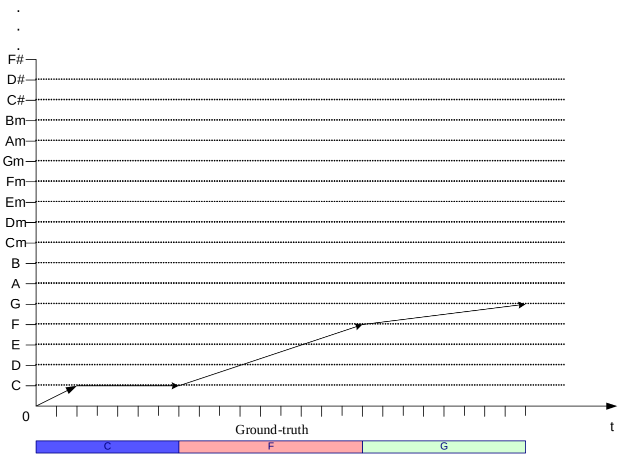

Figure 3. The best path in the lattice.

The Viterbi decoding outputs the most likely sequence of hidden states called the Viterbi path. This path is used to derive the output labels for the first configuration of the submitted system. At the same time Viterbi decoding can produce a lattice. The lattice structure can be represented as a directed graph with nodes denoting chord hypotheses and arcs denoting chord transitions. Figure 2 shows an example of a lattice, when in Figure 3 the ground-truth path is indicated. It is likely that this path obtains high score during the lattice rescoring procedure. The path with the highest score is used to form the output labels for the second configuration of the submitted system.

The vocabulary of the suggested chord recognition system consists of 12 major chords, 12 minor chords and non-chord segments. In the system with language modeling a 4-gram FLM configuration was used as described in [13].

4. RESULTS AND DISCUSSION

The evaluation results of all submitted systems to the pretrained subtask are presented in Figure 4. KO1 and KO2 correspond to the baseline system configuration and the configuration with language modeling part. It is worth noting that in the baseline approach no statistical information about chord transitions was used. Transition probabilities between each chord pair are equal and classification is based solely on acoustic features. Including language modeling showed a slight increase in performance.

5. REFERENCES

[1] Heng Tze Cheng, Yi-Hsuan Yang, Yu-Ching Lin, I-Bin Liao, and Homer H. Chen. Automatic chord recognition for music classification and retrieval. In *ICME*, pages 1505–1508. IEEE, 2008.

[2] Kyogu Lee. Identifying cover songs from audio using harmonic representation. In *MIREX task on Audio Cover Song Identification*, 2006.

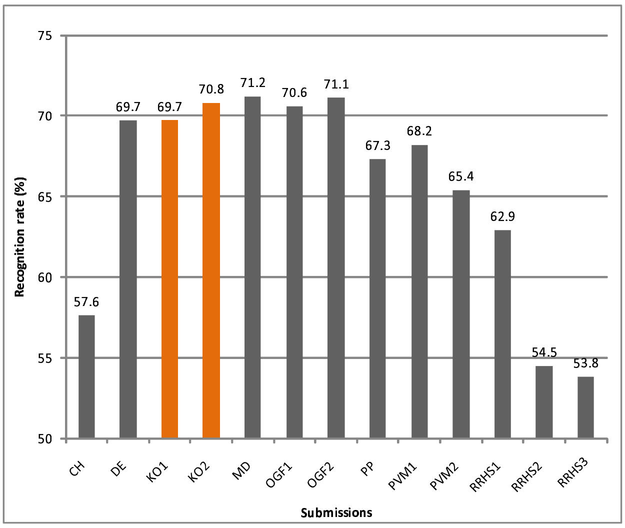

Figure 4. Evaluation results of different submissions.

[3] Juan P. Bello and Jeremy Pickens. A robust mid-level representation for harmonic content in music signal. In *Proceedings of the 2005 ISMIR Conference*, London, 2005.

[4] H. Papadopoulos and G. Peeters. Local key estimation based on harmonic and metric structures. In *Proceedings of DAFX*, Como, Italy, 2009.

[5] Takuya Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In *Proceedings of the International Computer Music Conference*, Beijing, 1999.

[6] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent hmms trained on synthesized audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), february 2008.

[7] Yushi Ueda, Yuki Uchiyama, Takuya Nishimoto, Nobutaka Ono, and Shigeki Sagayama. Hmm-based approach for automatic chord detection using refined acoustic features. In *Proc. ICASSP*, 2010.

[8] C. Harte and M. Sandler. Automatic chord identification using a quantized chromagram. In *Proceedings of the Audio Engineering Society*, Spain, 2005.

[9] Matthias Mauch and Simon Dixon. A discrete mixture model for chord labelling. In *Proceedings of the 2008 ISMIR Conference*, Philadelphia, 2008.

[10] H. Papadopoulos and G. Peeters. Simultaneous estimation of chord progression and downbeats from an audio file. In *Proc. ICASSP*, 2008.

[11] J. Stephen Downie. The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. *Acoustical Science and Technology. Available at: http://dx.doi.org/10.1250/ast.29.247*, 29(4):247–255, 2010.

[12] Daniel P.W. Ellis. The 2009 labrosa pretrained audio chord recognition system. In *Avalable at http://www.music-ir.org/mirex/2009/results/abs/DE.pdf*, 2009.

[13] M. Khadkevich and M. Omologo. Use of hidden markov models and factored language models for automatic chord recognition. In *Proceedings of the 2009 ISMIR Conference*, Kobe, Japan, 2009.

[14] Matthias Mauch and Simon Dixon. Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 2009.

[15] Yves Grenier Laurent Oudre and Cédric Févotte. Template-based chord recognition : Influence of the chord types. In *ISMIR*, 2009.

[16] J. Bilmes and K. Kirchoff. Factored language models and generalized parallel backoff. In *HLT-NAACL*, 2003.

[17] K. Kirchhoff, D. Vergyri, K. Duh, J. Bilmes, and A. Stolcke. Morphology-based language modeling for arabic speech recognition. In *Computer, Speech and Language*, 2006.

[18] S. Dixon. Onset detection revisited. In *Proceedings of DAFX*, McGill, Montreal, Canada, 2006.

[19] M. Khadkevich and M. Omologo. Phase-change based tuning for automatic chord recognition. In *Proceedings of DAFX*, Como, Italy, 2009.