Journal of Korea Multimedia Society Vol. 20, No. 9, September 2017(pp. 1502-1508)
https://doi.org/10.9717/kmms.2017.20.9.1502

Music Key Identification using Chroma Features and Hidden Markov Models

Pamela Kanyange†, Bong-Kee Sin††

ABSTRACT

A musical key is a fundamental concept in Western music theory. It is a collective characterization of pitches and chords that together create a musical perception of the entire piece. It is based on a group of pitches in a scale with which a music is constructed. Each key specifies the set of seven primary chromatic notes that are used out of the twelve possible notes. This paper presents a method that identifies the key of a song using Hidden Markov Models given a sequence of chroma features. Given an input song, a sequence of chroma features are computed. It is then classified into one of the 24 keys using a discrete Hidden Markov Models. The proposed method can help musicians and disc-jockeys in mixing a segment of tracks to create a medley. When tested on 120 songs, the success rate of the music key identification reached around 87.5%.

Key words: Music Key, Hidden Markov Model, Chroma Features, Machine Learning

1. INTRODUCTION

Music key is an essential feature in music analysis governing the entire music [1]. However, identifying the key of a song is very tough even for human. In music theory, we have 24 different keys that include 12 major and 12 minor keys. Each key consist of a progression of seven different pitches. To grasp the concept of keys, let's take an example of the key of "C major", where a song revolves around the seven notes of the C major scale C, D, E, F, G, A, and B. This means that the fundamental notes making up a song's melody, the majority of the chords in the songs must be derived primarily from the group of seven notes.

A challenging problem in music key identification is that some keys are very similar to each other and very difficult to discriminate, if not impossible. Fig. 1 shows the circle of fifths giving the relationship among the 12 pitches of the chromatic scales to their corresponding keys major and minor [4]. And the majority of the misclassifications are caused by the difficulty in distinguishing the keys closely related in the diagram. Manually labeling the key of many songs is an extremely time-consuming and tedious task.

The problem of automatic key identification is not new. Among the variety of methods developed to date, the Krumhansl-Schmuckler key-finding model is the most popular tool where the distribution of pitches in a piece is compared with the ideal distribution or "key profile" of each key [5] [6, 7]. A. Shenoy et al. proposed the idea of combining Chroma based frequency analysis and music knowledge of rhythm structure and chord change patterns followed by rule-based inference [8].

※ Corresponding Author : Bong-Kee Sin, Address: (48513) 45, Yongso-ro, Nam-Gu. Busan, Korea, TEL : +82-51-629-6256, FAX : +82-51-629-6230, E-mail : bkshin@pknu.ac.kr
Receipt date : May 4, 2017, Revision date : Jul. 19, 2017
Approval date : Aug. 9, 2017

† Department of IT Convergence and Applications Engineering, Pukyong National University (E-mail : kayparme@yahoo.fr)
†† Department of IT Convergence and Applications Engineering, Pukyong National University
※ This work was supported by a Research Grant of Pukyong National University (2016)

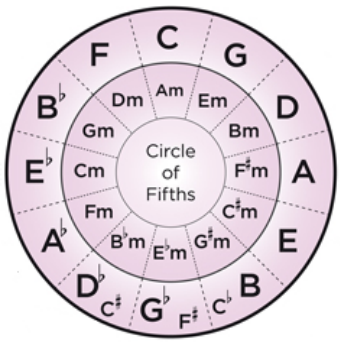

Fig. 1. The circle of fifths [4].

For [10], they presented a method for estimating the root of diatonic scale and the key directly from acoustic signals (waveform) of popular and classical music. They proposed a method to extract pitch profile features from the audio signal, which characterizes the tone distribution in the music. The diatonic scale root and key are estimated based on the extracted pitch profile by using a tone clustering algorithm and utilizing the tone structure of keys.

In another research that developed by Chew [10], They proposed a Boundary Search Algorithm (BSA) for determining points of modulation in a piece of music using a geometric model for tonality called the Spiral Array. For a given number of key changes, the computational complexity of the algorithm is polynomial in the number of pitch events.

There is another approach introduced by Xinquan and Alexander [11]. In this work, they investigate Deep Networks (DNs) for learning high-level and more representative features in the context of chord detection, effectively replacing the widely used pitch chroma intermediate representation.

This paper proposes a method for finding the music key of a song by extracting chroma features and determining the key via Hidden Markov Models. The study has been conducted on several different genres of music by many different composers. One limitation of this research is that it's very difficult to obtain a training dataset with correct labels. We aim to show through our analysis that this model returns consistent results through experiments. The organization of the proposed approach is illustrated in Fig. 2.

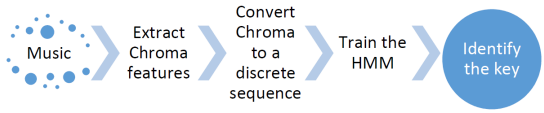

Fig. 2. The architecture of the music key identification system.

2. CHROMA FEATURES

Chroma is an audio music feature developed for music analysis based on pitches and a pitch class is a set of all pitches sharing the same chroma [2]. In the music discipline, the term chroma relates to the twelve different pitch classes used in Western music notation: C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B. In chroma, all notes that have the same chroma value belong to the same pitch class. For instance, the pitch class that corresponds to the chroma C will consist of the set of C0, C1, C2, C3, and so on, each from different octaves. Each chroma feature represents the intensity associated with each of the 12 semitones within one octave, but all octaves share the chroma values. Let us consider a piano keyboard, the chroma C refers to all the C notes irrespective of the octave, high C or low C.

The advantage of using chroma feature is that they capture harmonic and melodic characteristics of music while being robust to variability in timbre and instrumentation [2]. Every pitch that we perceive corresponds to a particular frequency of a sinusoid in a sound signal. That is why the chroma feature became a major tool for processing and analyzing music data. Chroma features are derived from the energy found within a given frequency range in short-time spectral representations of audio signals extracted on a frame-by-frame basis. They segment the audio signal into narrow time intervals and take the Fourier transform of each

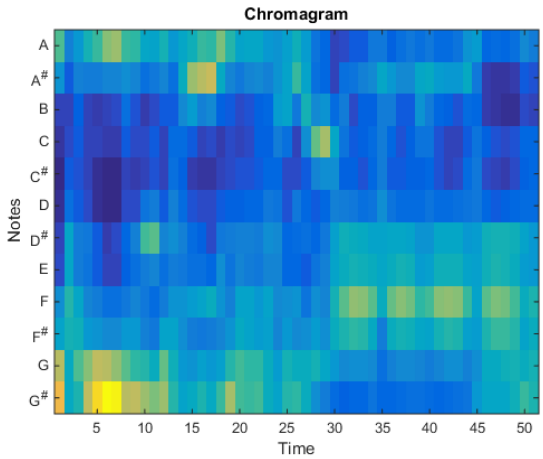

Fig. 3. A Chromagram representation.

segment. Consider a signal $f(t)$ to be analyzed at time t we compute the short-time Fourier transform (STFT) for each window W :

$$STFT_f^u(t',u) = \int [f(t)\, W(t-t')]e^{-j2\pi ut}dt \quad (1)$$

The main idea of chroma features is to combine all spectral information (1) that relates to a given pitch class into a single coefficient. In other words, it's the sum of the spectral energy overall.

$$C_f(b) = \sum_{z=0}^{Z-1} |X(b+z\beta)| \quad (2)$$

where X is the log-frequency spectrum, z the octave index $\in [0, Z-1]$, Z the number of octaves and b the pitch class (chroma) index $\in [0, \beta-1]$ where β is the number of bins per octave.

A chromagram is a 12 by M matrix which contains a sequence of measurements about of the strength of the 12 possible notes in each of the M spectrogram windows, across all octaves.

3. HIDDEN MARKOV MODEL

Hidden Markov Model is a very popular tool for modeling time series data [12]. It describes the probabilistic generation of a sequence of observations. It also provides a way of computing the joint probability of a set of hidden and observed discrete random variables.

3.1 Model training

HMM is a popular stochastic modeling tool comprising three sets of parameters as denoted in the triple $\lambda = (A, B, \pi)$, where A is the transition matrix of $A_{ij} =$ Pr(transition from state i to state j), B the emission matrix of probabilities as $B_{jk} =$ Pr(emission of symbol k from state j) and as the state prior $\pi = (\pi_i)_i$. In the training phase, the model parameters are estimated to maximize the likelihood of the model given observed sequences. Baum-Welch algorithm also known as Expectation Maximization is used to find the maximum likelihood parameters [13]. Given a sequence, we compute the posterior estimates of various hidden variables in the E-step using the forward-backward algorithm. Based on these quantities, we optimize the model parameters in the M-step. The topology of an HMM also matters for improving performance. In this research we employed the ergodic topology for all of the 24 key HMMs, where any transitions are possible.

3.2 Recognition

In the recognition phase, we employ the Bayes classifier with equal class priors. Thus the classifier is simply likelihood-based, given a test sequence Y, and we compute the log likelihood of the model λ_k, as follows:

$$\log P(Y|\lambda_k) = \log \sum_X P(Y, X|\lambda_k), \; k = 1,...,24. \quad (3)$$

where X represents an arbitrary Markov chain $X_1, X_2, ..., X_T$. This can be efficiently computed using the forward algorithm on the Viterbi algorithm. Then, the model that returns the highest log-likelihood is selected as the one representing the key of the song.

$$key = \arg\max_k \log P(Y|\lambda_k) \quad (4)$$

4. EXPERIMENTS

The proposed method has been implemented in MATLAB using the HMM toolbox by Kevin Murphy [14]. To train the key HMMs models, we collected training samples of 10 songs for each key and additional 5 songs for each key to test the model performance.

Classes	Train sets	Test sets
24	240	120

For both training and testing, we first extract the chroma features from each song and created a sequence of indices to the maximum element in each column of the 12-dimensional chromagram. A sample result is shown in Fig. 4.

We have tested HMMs with varying number of states (5, 10, 15, 20, 25, 30, 50 states), then conducted a series of tests on the models to find the optimal number of states. According to the test result as shown in Fig. 5, with 25 states was found to be the best choice.

The result of this analysis shows that the proposed technique is reliable with an accuracy up to 87.5% and the proposed method is not limited to any style of composition and audio styles. See Fig. 8 for a detailed clarification behavior.

In this experiment, the prototype system based on the proposed method detected the keys correctly, 105 out of the 120 test songs.

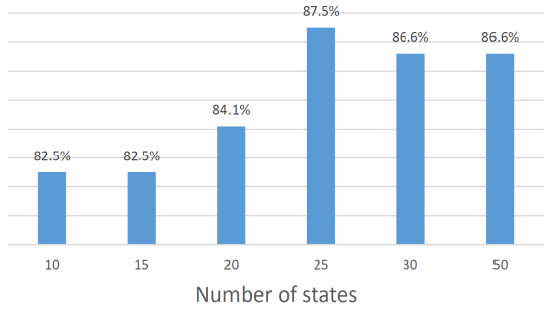

Fig. 5. HMM performance with different number of states.

Table 1.Classification accuracy from different models

Model	Accuracy
Markov Chain	81.6%
DHMM	87.5%
CHMM	77.5%

5. PERFORMANCE COMPARISON

In the final set of experiments, we compared the classification performance of four different models, i.e. histograms, Markov Chains, discrete and continuous HMMs.

Table 1 gives a summary of the result. Despite the simplicity of DHMM compared to the CHMM, the former exhibited the best result. And even the simple Markov chains fared better than the CHMMs which are believed to need more data to work better. Based on the set of test results re-

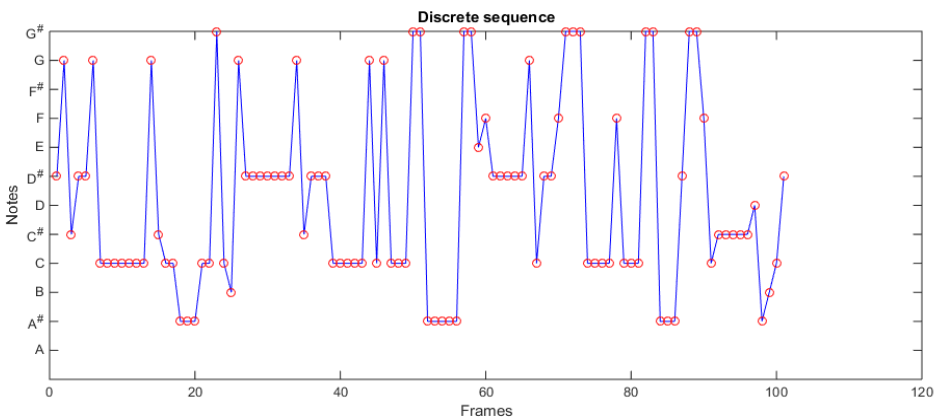

Fig. 4. Discrete sequence of pitch codes derived from a current chromagram.

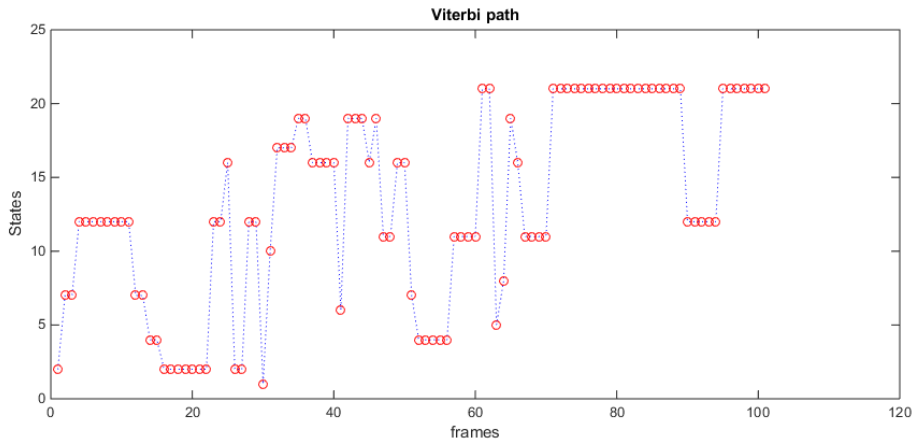

Fig. 6. The Viterbi path of an HMM with 25 states given a song as represented as the sequence in Fig 4. This shows us the most probable path found using the Viterbi algorithm.

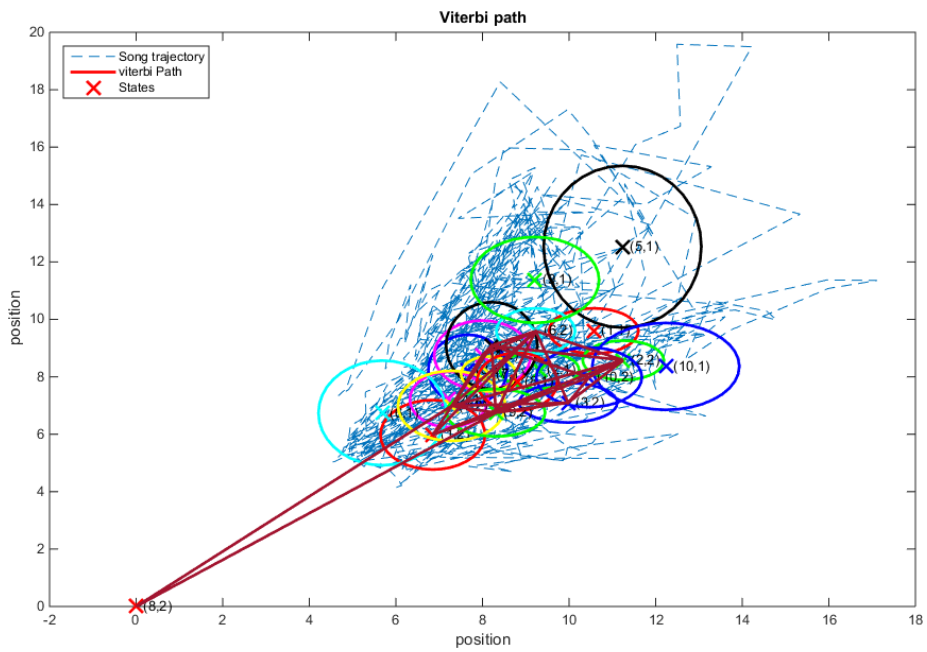

Fig. 7. The broken polyline represents a song trajectory in feature space. The ellipses denote HMM state Gaussians. The central solid lines correspond to an annotation made by a key HMM using the Viterbi algorithm. Although messy, it shows that the state Gaussians appropriately models the input signals.

viewed this far, we can conclude that discrete HMM is a more powerful approach to this problem and under the current experiment setting. But we also want to mention that continuous HMM should have a potential for improvement given sufficient training set. Continuous HMM is an extension of DHMM. It makes the HMM work with continuous observations instead of only allowing discrete sequence. In CHMM, the parameter B, the state probability of the model cannot be represented as a simple matrix of point probabilities but rather as a complete probability density function with continuous observation.

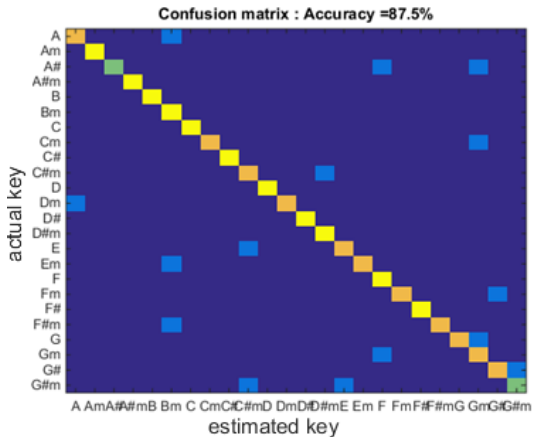

Fig. 8. The Confusion matrix detailing the accuracy of the estimated keys are compared to the known actual keys. Overall 87.5% of the songs were correctly classified.

6. CONCLUSIONS

This paper presents a solution to music key identification. The proposed method first extracts the chroma features from a song and converts the chromagram to a sequence of discrete symbols which will be used to train and test a discrete HMM. In our experiments, we collected 240 songs of various genres from different musicians. Performance evaluation with 120 songs marked a classification accuracy of 87.5%.

Future direction of research lies in the optimization of the training method by changing the meta-parameters of the model and increasing the datasets. We are also planning to make an automated disk-jockey software by joining together other music processing algorithms we already made in our previous researches.

REFERENCE

[1] J.D. White, *The Analysis of Music,* Prentice-Hall, Englewood Cliffs, N.J., 1976.

[2] A.M. Wisnu, M. Carmadi, S.P. Ary, and B.K. Sin, "Design of Music Learning Assistant Based On Audio Music and Music Score Recognition," *Journal of Korea Multimedia Society*, Vol. 19, No. 5, pp. 826-836, 2016.

[3] C.M. Grinstead and J.L. Snell, *Introduction to Probability*, American Mathematical Society, 2003.

[4] Circle of Fifths, Wikipedia, https://en.wikipedia.org/wiki/Circle_of_fifths, (accessed May, 01, 2017).

[5] Y. Zhu and M.S. Kankanhalli, "Precise Pitch Profile Feature Extraction from Musical Audio for Key Detection," *IEEE Transactions on Multimedia*, Vol. 8, No. 3, pp. 575-584, 2006.

[6] S. Pauws, "Musical Key Extraction from Audio," *Proceeding of the Fifth International Conference on Music Informational Retrieval*, vol.4, pp.66-69, 2004.

[7] C.L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, 1990.

[8] A. Shenoy, "Key Determination of Acoustic Musical Signals," *Proceeding of IEEE International Conference on Multimedia and Expo*, pp.1771-1744, 2004.

[9] E. Chew, "The Spiral Array: An Algorithm for Determining Key Boundaries," *Proceedings of the 2nd International Conference on Music and Artificial Intelligence*, pp.18-31, 2002.

[10] Y. Zhu, M.S. Kankanhalli, and S. Gao, "Music Key Detection for Musical Audio," *Proceeding of 11th International Mathematical Modeling Challenge*, pp. 30-37, 2005.

[11] Z. Xinquan and L. Alexander, "Chord Detection Using Deep Learning," *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pp. 26-30, 2015.

[12] R.N. Shepard, "Circularity in Judgments of Relative Pitch," *The Journal of the Acoustical Society of America*, Vol. 36, pp. 2346, 1964.

[13] L.R. Rabiner, "Hidden Markov Models for Speech Recognition," *Technometrics: American Statistical Association*, Vol. 33, No. 3, pp. 251-

272, 1991.

[14] K. Murphy, Hidden Markov Model (HMM) Toolbox for Matlab, https://www.cs. ubc.-ca/~murphyk/Software/HMM/hmm.html (accessed May, 01, 2017).

Kanyange Pamela

In 2012, she graduated from Symbiosis International University with a degree in computer applications.
From 2013 to 2015, she worked in a telecom company based in Burundi as a billing engineer.
In 2015, she started her master degree in IT Convertgence and Applications Engineering at Pukyong National University in Korea. A great lover of music, she oriented her research on music applications using machine learning tools.

Bong-Kee Sin

1985, Bachelor degree from the Department of Mineral and Petroleum Engineering, Seoul National University.
1987, Master degree from the department of Computer Science, Korea Advanced Institute of Science and Technology.
1995, PhD from the Department of Computer Science, KAIST.
1987~1999, Senior Researcher, SW Research Laboratories, Korea Telecom.
1999~present, Professor in the Department of IT Convergence and Applications Engineering, Pukyong National University.
Research interest: pattern recognition, machine learning, computer vision and artificial intelligence.