Proceedings of the International MultiConference of Engineers and Computer Scientists 2016 Vol I,
IMECS 2016, March 16 - 18, 2016, Hong Kong

Method for Extraction of Chord Progressions from Pictures by 2D FFT

Hidefumi Kawakatsu*

***Abstract*—This study proposes a method for generating chord progressions from pictures applying the two-dimensional fast Fourier transform (2D FFT) algorithm. The picture is first divided into some squares, and then the image in each square is decomposed into a plurality of frequency components. We consider the case where the frequencies obtained from each square are transformed into a chord through combining two kinds of intervals: (a) the difference between the frequency components and (b) the difference between pitches in music theory. Some examples are provided to illustrate the proposed method.**

***Index Terms*—fast Fourier transform, chord progression, interval, music extraction**

I. INTRODUCTION

Sightseeing destinations have been introduced on brochures or web sites with attractive photographs or illustrations. Music as well as pictures can be an effective medium to communicate their appeals to tourists. This study proposes a method for generating chord progressions from pictures applying the two-dimensional fast Fourier transform (2D FFT) algorithm. The method proposed here may be available for displaying, on the web sites, the pictures along with the music generated from them.

Several researchers have analyzed art works such as music and paintings by means of fluctuation analysis. They have shown that the art works display regularities in scaling properties and long-range correlations[1], [2], [3], [4]. They indicated that pleasant music for humans displays a behavior similar to $1/f$ noise. Dagduga et al.[3] considered the music score as a sequence of integer numbers, and then quantified long-range correlations in a Mozart's music score by using DFA (detrended fluctuation analysis). Ro and Kwon[4] investigated the behavior of $1/f$ noise in songs which belong to seven different genres, and showed their $1/f$ analysis of songs in the region below 20 Hz can classify them into each genre of music. For visual arts, Rodriguez et al. [5] have clarified the existence of the factuality and scaling properties in Pollock's drip paintings. They suggested that paintings which contain $1/f$-noise structures can also stimulate the perception of pleasant. They applied two-dimensional DFA method to analyze the gray-scaled images obtained from paintings. The DFA algorithm for two dimensions has simpler structure, but its computation time becomes greater than that of the 2D FFT.

In our previous study, we have proposed methods for finding pleasant photographs of tourist destinations and for extracting music from pictures by applying the 2D DFA and 2D FFT[6]. The method for extraction of music proposed there, however, generated the music without considering musical theory.

Manuscript received December 31, 2015; revised January 26, 2016.
H. Kawakatsu is with Department of Economics & Information Science, Onomichi City University, 1600-2 Hisayamadacho, Onomichi 722-8506 Japan (corresponding author to provide phone: +81-848-22-8312 (ex.616); fax: +81-848-22-5460; e-mail: kawakatsu@onomichi-u.ac.jp).

This study proposes a method for generating chord progressions from pictures applying the two-dimensional fast Fourier transform (2D FFT) algorithm. The picture is first divided into some squares, and then the image in each square is decomposed into a plurality of frequency components. We consider the case where the frequencies obtained from each square are transformed into a chord through combining two kinds of intervals: (a) the difference between the frequency components and (b) the difference between pitches in music theory. Some examples are provided to illustrate the proposed method.

II. Spectral Analysis using 2D FFT

This section provides the method for obtaining the sequence of frequency components from the picture by applying the two-dimensional FFT (2D FFT) to reflect the features on the surface of the image, which will be used for extraction of chords from the picture in Section III.

The color images are transformed into gray scale images as

$$G = 0.299R + 0.587G + 0.114B. \qquad (1)$$

The size of these images is normalized to N (vertical) $\times M$ (horizontal) pixels, where N is determined so that the aspect ratio of the image can be maintained, and M is set to be 2^k $(k = 2, 3, \cdots)$. Let $G_n(m)$ $(m = 1, 2, \cdots, M,$ $n = 1, 2, \cdots, N,$ $G_n(m) = 0, 1, \cdots, 255)$ be the brightness (gray scale level) at the coordinates (m, n) of the gray scale image.

1) Divide the image into L squares of length of a side of h $(h = 2^k)$, where L $(= L_1 \times L_2)$ denotes the number of squares, $L_1 = \lfloor M/h \rfloor$ and $L_2 = \lfloor N/h \rfloor$.
2) $l = 1$.
3) Set $g_l(u, v) = G_{y+v}(x + u)$ $(l = 1, 2, \cdots, L,$ $u = 1, 2, \cdots, h,$ $v = 1, 2, \cdots, h)$, $x = [(l-1) \bmod h] \times h$, $y = [\lfloor (l-1)/3 \rfloor \times h$.
4) Compute the power spectral density $P_l(u, v)$ from $g_l(u, v)$ using two-dimensional FFT (2D FFT), and calculate the average of $P_l(u, v)$, which is given by
$$\bar{P}_l(r) = \frac{1}{\pi} \sum_{\theta=1}^{\pi} P_l(r, \theta), \qquad (2)$$
where $r = \sqrt{u^2 + v^2}$.
5) Find the frequencies $r_{(i)}$ $(i = 1, 2, \cdots, \tau)$ corresponding to the local maxima for the average values of spectral densities $\bar{P}_l(r)$ in Eq. (2), where $r_{(1)} < r_{(2)} < \cdots < r_{(\tau)}$ and $3 \le \tau < h/2$. Let $\boldsymbol{x}_l = (x_1, x_2, \cdots, x_\tau) = (r_{(1)}, r_{(2)}, \cdots, r_{(\tau)})$.
6) If $l < L$, increment l by one and go to Step 3).

TABLE I
MAIN INTERVALS

Number of semitones	Intervals
0	P1
1	m2
2	M2
3	m3
4	M3
5	P4
6	dim5
7	P5
8	aug5
9	M6 or dim7
10	m7
11	M7
12	P8

III. Method for Extraction of Chords

Musical chords have the following properties:

(i) The quality of a compound interval is determined by the quality of the simple interval on which it is based.

(ii) Chords with same notes can be classified into the same chord. For instance, Am7 (with the notes ACEG) is the same chord as C6 (with the notes CEGA).

Let $A = \{1, 2, \cdots, 108\}$ and $B = \{1, 2, \cdots, 12\}$. We assign an integer $s \in A$ to each musical note of the scale such that middle C (C4) is assigned 49, the note just above (C♯4 or D♭4) is 50 and the musical rest is 0. We refer to s as "note number".

In this case, the above property (i) leads to the following property (iii):

(iii) The span of semitones, denoted by $d(s_1, s_2)$, to derive the interval between two pitches can be given by

$$d(s_1, s_2) = |f(s_2) - f(s_1)|, \quad (s_1, s_2 \in A,\ s_1 \leq s_2) \tag{3}$$

where $f : A \to B$ is a surjection since $f(A) = B$, which is expressed by

$$f(s) = [(s-1) \mod 12] + 1. \tag{4}$$

We propose the method for extraction of chord progressions on the basis of the above observations. The method consists of following six steps:

1) $l = 1$.
2) Find the $\boldsymbol{x}_l$, which is given by Step 5) in Section II. Each component x_i can be transformed into $y_i \in B$ by the equation $y_i = f(x_i)$ in Eq. (4) since the compound interval is determined by the simple interval on which it is based, as explained in Property (i). Assume that y_j expresses a temporarily root of the chord, which is referred to as "interim root" in this study. Then $\boldsymbol{y}_l^{(j)}$ with y_j as the interim root is given by an "inversion of intervals", i.e., by shifting y_i $(i < j)$ up one octave:

$$\begin{aligned}\boldsymbol{y}_l^{(j)} &= \left(y_{(1)}^{(j)}, y_{(2)}^{(j)}, \cdots, y_{(\tau)}^{(j)}\right) \\ &= (y_j, y_{j+1}, \cdots, y_\tau, \\ &\qquad y_1 + 12, y_2 + 12, \cdots, y_{j-1} + 12).\end{aligned} \tag{5}$$

TABLE II
EXAMPLE OF TRAINING SAMPLES

k	$\boldsymbol{Q}_k$	C_k
1	$(0,1,0,0,1,0,0,0,0,0)$	□
2	$(0,1,0,0,0,1,0,0,0,0)$	□$^{+5}$
3	$(0,0,1,0,1,0,0,0,0,0)$	□sus4
4	$(1,0,0,0,1,0,0,0,0,0)$	□m
5	$(1,0,0,1,0,0,0,0,0,0)$	□m^{-5}
6	$(1,0,0,0,0,1,0,0,0,0)$	□m^{+5}
7	$(0,1,0,0,1,0,1,0,0,0)$	□6
8	$(1,0,0,0,1,0,1,0,0,0)$	□m6
9	$(0,1,0,0,1,0,0,0,0,1)$	□M7
10	$(1,0,0,0,1,0,0,0,0,1)$	□mM7
11	$(1,0,0,1,0,0,0,0,0,1)$	□dimM7
12	$(0,1,0,0,1,0,0,0,1,0)$	□7
13	$(0,1,0,0,0,1,0,0,1,0)$	□7^{+5}
14	$(0,1,0,1,0,0,0,0,1,0)$	□7
15	$(0,0,1,0,1,0,0,0,1,0)$	□7sus4
16	$(1,0,0,0,1,0,0,0,1,0)$	□m7
17	$(1,0,0,1,0,0,0,0,1,0)$	□m7^{-5}
18	$(1,0,0,1,0,0,0,1,0,0)$	□dim7

3) Measure the intervals between the interim root $y_{(1)}^{(j)}$ and $y_{(t)}^{(j)}$ $(t = 2, 3, \ldots, \tau)$ through calculation of the difference of semitones, $d(y_{(1)}^{(j)}, y_{(t)}^{(j)})$, between them, which can be derived by Eq. (3). Table I shows the intervals mainly used in this study corresponding to the number of semitones.

We here introduce some more additional notation as follows:

$\eta(y_{(t)}^{(j)})$**:** interval between $y_{(1)}^{(j)}$ and $y_{(t)}^{(j)}$ $(t = 2, 3, \ldots, \tau)$.

$\boldsymbol{z}^{(j)}$**:** feature vector of "unknown pattern", where $\boldsymbol{z}^{(j)} = (z_1^{(j)}, z_2^{(j)}, \cdots, z_{10}^{(j)})$ $(z_b^{(j)} \in \{0, 1\})$.

In this case, the first component $z_1^{(j)}$ in $\boldsymbol{z}^{(j)}$ is set to be 1 if $\eta(y_{(t)}^{(j)})$ is "m1", otherwise $z_1^{(j)}$ becomes 0. The remaining components, $z_2^{(j)}, z_3^{(j)}, \cdots, z_{10}^{(j)}$, in $\boldsymbol{z}^{(j)}$, correspond to "M3", "P4", "dim5", "P5", "aug5", "M6", "dim7", "m7", "M7", respectively. It should be noted that the number of semitones of "M6" and that of "dim7" take the same value of 9. Therefore, $z_7^{(j)}$ (corresponding to M6) and $z_8^{(j)}$ (corresponding to dim7) are respectively set to be 0 and 1 if z_1 (m3) = 1 and z_4 (dim5) = 1, otherwise $(z_7^{(j)}, z_8^{(j)}) = (1, 0)$.

4) Calculate the degree of similarity between $\boldsymbol{y}_l^{(j)}$ and an actual musical chords by means of deriving the simple matching coefficient between them. Table II shows training samples, where the square (□) represents a root of the chord. Then determine $C_k = C_{k^*}^{(j^*)}$ $(= C_l^*)$ and $\boldsymbol{y}_l^{(j)} = \boldsymbol{y}_l^{(j^*)}$ $(= \boldsymbol{y}_l^*)$ through finding $(j, k) = (j^*, k^*)$, which maximizes $S_k^{(j)}$ for all j and k, where

$$S_k^{(j)} = \sum_{b=1}^{10} \frac{(z_b^{(j)} \leftrightarrow q_{k,b})}{10}. \tag{6}$$

5) Determine the root of the chord extracted in the previous step as follows:

We suppose that the chord of I is most frequently

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Proceedings of the International MultiConference of Engineers and Computer Scientists 2016 Vol I,
IMECS 2016, March 16 - 18, 2016, Hong Kong

TABLE III
HARMONIC FUNCTIONS IN C MAJOR

Harmonic functions	Main chord	Substitute chord
Tonic	I(C)	III(Em), VI(Am)
Dominant	V(G)	[VII (Bm^{-5})]
Sub-dominant	IV(F)	II(Dm)

used in the chord progression. Then the most frequent value, denoted by y_{mo}, of the first component in $\boldsymbol{y}_l^*$ for $l = 1, 2, \cdots, L$ can be estimated as the root of I. For instance, if the key of the objective chord progression is C, y_{mo} is considered as the note of C. In this case, the notes (note numbers) in $\boldsymbol{y}_l^*$ can be transformed by $y_{(t)}^* + d$, where $d = 49 - y_{mo}$ $(t = 1, 2, \cdots, \tau)$ since 49 is assigned as C4 as mentioned above.
In the case where other keys are used, the root of I can be estimated without loss of generality since the keynote of its scale can be used for this estimation.

6) If $l < L$, increment l by one and go to Step 2).

IV. Examples of the Proposed Method

This section presents examples for the extraction of chord progressions from some pictures to illustrate the proposed method in this study.

We focus on the following cases for simplicity:

(i) The chord progression is generated in C major.

(ii) The chords obtained from Step 4) in Section III are arranged so that the chord progression may include cadences (T-S-T, T-D-T and T-S-D-T) in order of their appearance. Table III summarizes chord names in C major corresponding to harmonic functions (T, D and S), where T, D and S respectively denote Tonic, Dominant and Sub-dominant. For instance, if the sequence of (C, C, F, G) is extracted from the procedure proposed in Section III, the chord progression of C-F-G-C (T-S-D-T) is then obtained.

(iii) The chords that the root is a natural tone are chosen from the chords extracted from the picture to generate a chord progression in C major. This signifies that the length of the generated chord progression depends on the number of extracted chords with the root of the natural tone.

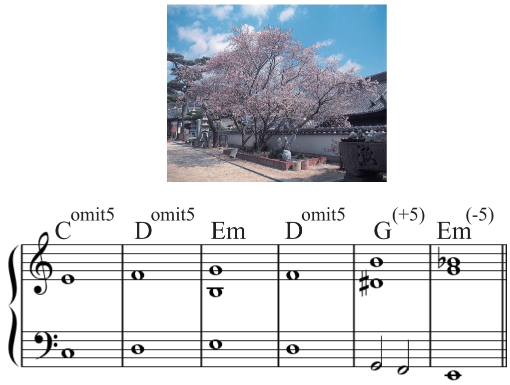
(iv) The chords can be inverted to be included the strong cadence, the semitone descending (or ascending) and/or parallel motion in the base line (bass line) as many as possible. The note of M3 is inserted just behind the chords of sus4, if any, which can be expected to resolve to a major chord.

We also consider the case where the picture divided into 16 squares, which indicates that 16 chords are first extracted from a picture.

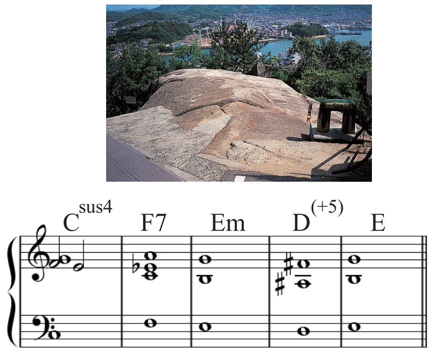
The procedure proposed in Section III first generates the following chords from the picture in Fig. (1)(a):

D(omit5), C(omit5), C♯m$^{(+5)}$, Em, F♯sus4 C♯m$^{(-5)}$, C♯m, C♯m, Em$^{(-5)}$, F♯m$^{(+5)}$, Dm(omit5) C C♯$^{(+5)}$, C♯m(omit5) D♯ and G$^{(+5)}$.

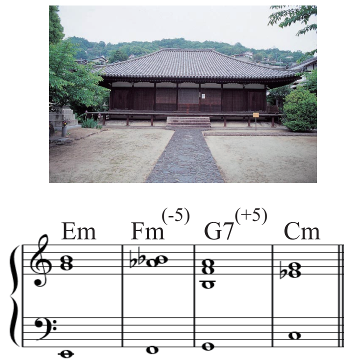
By selecting the chords with the root of the natural tones from this sequence, we then obtain the sequence which contains:



(a) Example 1



(b) Example 2



(c) Example 3

Fig. 1. Examples

D(omit5), C(omit5), Em, Em$^{(-5)}$, Dm(omit5), C and G$^{(+5)}$.

By sorting these so as to satisfy the conditions (ii) and (iv), the chord progression is generated as shown in Fig. (1)(a).

Figure (1)(b) and (c) also reveal the different pictures and the chord progressions generated from these pictures.

V. Conclusion

This study have developed a method for generating chord progressions from pictures applying the two-dimensional fast Fourier transform (2D FFT) algorithm. The picture is first divided into some squares, and then the image in each square is decomposed into a plurality of frequency components. We have considered the case where the frequencies in each square are transformed into a chord through combining two kinds of intervals: (a) the difference between the frequency components and (b) the difference between pitches in music theory. We could extract chords which "embed" or "hidden" in the picture considering the inversion and the interval in musical theory. Some examples are provided to illustrate the proposed method.

This study extracted the chords by applying the 2D FFT, but by using the wavelet analysis, a wide variety of chord progressions can be generated from the picture. Taking account of such factors is an interesting extension.

REFERENCES

[1] R. F. Voss and J. Clarke, ""$1/f$ noise" in music: Music from $1/f$ noise," *J. Acoust. Soc. Am.*, vol. 63, no. 1, pp. 258–263, 1978.
[2] G. R. Jafari, P. Pedram, and L. Hedayatifar, "Long-range correlation and multifractality in bach's inventions pitches," *Journal of Statistical Mechanics: Theory and Experiment*, vol. P04012, 2007.
[3] L. Dagduga *et al.*, "Correlations in a mozart's music score (k-73x) with palindromic and upside-down structure," *Physica A*, vol. 383, no. 2, pp. 570–584, 2007.
[4] W. Ro and Y. Kwon, "$1/f$ noise analysis of songs in various genre of music," *Chaos Solitons & Fractals*, vol. 42, pp. 2305–2311, 2009.
[5] E. Rodriguez *et al.*, "$1/f$ noise structures in pollocks' drip paintings," *Physica A*, vol. 387, pp. 281–295, 2008.
[6] H. Kawakatsu, "Methods for evaluating pictures and extracting music by 2d dfa and 2d fft," *Procedia Computer Science*, vol. 60, pp. 834–840, 2015.

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)