

# Selection of Appropriate Methods for Classifying Pictures and Generating Chord Progressions to Introduce Tourist Destinations

Hidefumi Kawakatsu <sup>\*</sup>, Kosuke Kato <sup>†</sup>

## Abstract

This study provides the following two methods applying the 2D FFT (fast Fourier transform) algorithms: (1) a method for finding pleasant photographs of local tourist destinations based on the results of a preliminary experiment, and (2) a method for generating chord progressions from these photographs considering the 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 methods.

*Keywords:* 2D FFT, classification of picture, music extraction, interval

## 1 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 communicative medium [1]. This study attempts to develop a tool for appealing tourist attractions through analyzing the pictures and generating music from them. Results of the analyses and music generated from pictures differ depending on the time they were taken, and change with angles. The method proposed here may be available to introduce the tourist spots in real time with “degree of pleasantness of picture”, defined in Section 2, and with music generated from pictures on a web site. The tourists may also be able to measure themselves the pictures and generate music if the methods are implemented in a portable device such as a smart phone.

Mathematics and art works such as music have some vague sort of affinity, but several researchers have shown that the music displays regularities in scaling properties and long-range correlations[2, 3, 4, 5]. They indicated that pleasant music for humans displays a behavior similar to  $1/f$  noise. For visual arts, Rodriguez, Echeverria and Alvarez-Ramirez[6] 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 pleasantness. They applied two-dimensional DFA (detrended fluctuation

---

<sup>\*</sup> Onomichi City University, Hiroshima, Japan

<sup>†</sup> Hiroshima Institute of Technology, Hiroshima, Japan

analysis) method to analyze the gray-scaled images obtained from paintings. The DFA algorithm for two dimensions has simpler structure than the two-dimensional FFT (fast Fourier transform), but its computation time becomes greater than that of the one-dimensional case. Musha[7] has described the method evaluating paintings by using one-dimensional Fourier transform as follows: (i) compute the power spectral densities of brightness in the horizontal direction in each row, and then (ii) calculate the average values for the spectral densities in each column. His method may not detect a self-similarity of the surface of the image since it focused on the behavior of the brightness oscillation in the horizontal direction.

We have already proposed a method applying the 2D DFA to detect a self-similarity on the surface of the image, and then compared his and our methods[8]. Our previous study proposed the method for classifying the pictures by combining 1D and 2D FFTs instead of applying 2D DFA[9]. This study develops the method using only 2D FFT to simplify the procedure based on a preliminary experiment using test image data and subjective evaluation.

This study also provides a method for a chord progression from a picture simultaneously with classifying them. The picture is first divided into several squares, and then the image in each square is decomposed into a plurality of frequency components. We consider the case where the frequency components obtained from each region are transformed into a chord through connecting 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.

## 2 Classification of Pictures

The  $1/f^\alpha$  noise is defined in terms of the shape of its power spectral density  $P(f)$ , where  $f$  is the frequency. When the signal follows a scaling law, a power-law behavior for the power spectral density  $P(f)$  is observed:

$$P(f) \sim 1/f^\alpha, \quad (1)$$

where  $\alpha$  is called the scaling (spectral) exponent. The scaling exponent  $\alpha$  is computed as the slope of the plot  $\{\log(f)$  versus  $\log(P(f))\}$ . The sequence, denoted by  $u(i)$ , can be classified as follows: (a) If  $\alpha = 0$ , the sequence  $u(i)$  is non-correlated (i.e. white noise); (b) If  $\alpha = -1$ , the correlation of the sequence is the same of  $1/f$  noise; and (c) If  $\alpha = -2$ , the sequence behaves like Brownian motion (random walk)[2, 10].

We define ‘‘pleasant photograph’’ as the photograph containing  $1/f$  components, and then propose a procedure for finding the pleasant photographs using the fast Fourier transform (FFT) algorithm. The color images are transformed into gray scale images by using a simple linear combination of the original RGB channels. The size of these images is normalized to  $M$  (horizontal)  $\times N$  (vertical) pixels, where  $M = 2^k$  ( $k = 2, 3, \dots$ ) and  $N$  is determined so that the aspect ratio of the image can be maintained. Let  $G_n(m)$  ( $m = 1, 2, \dots, M$ ,  $n = 1, 2, \dots, N$ ,  $G_n(m) = 0, 1, \dots, 255$ ) be the brightness (gray scale level) at the coordinates  $(m, n)$  of the gray scale image.

This study defines  $\alpha_d$  ( $d = 1, 2$ ) as the scaling (spectral) exponent obtained by the following methods using  $d$ -dimensional FFT. We subjectively evaluate 300 pictures based on interviews with students, professors of arts and civils. On the bases of this result, the pictures can be classified into one of the following three cases: (a) If  $\alpha \geq -0.87$ , the image is classified into ‘‘Random’’ (or ‘‘White’’): (b) If  $-0.87 \leq \alpha < -1.50$ , the image is classified

into “ $1/f$  noise” (or “Pink”): (c) If  $\alpha < -1.50$ , the image is classified into “Brownian” (or “Red”).

### 2.1 Method using 1D FFT (Method 1)

We have already proposed the method for classifying the pictures applying the one-dimensional FFT algorithm[11], which is referred to as “Method 1” in this study. The scaling exponent  $\alpha_1$  defined above can be calculated by Method 1.

### 2.2 Method using 2D FFT (Method 2)

Our previous study[9] has proposed the method for classifying the pictures using the two-dimensional FFT algorithm, which is referred to as “Method 2” in this study. The procedure consists of the following six steps:

**Step 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  $L_2 = \lfloor N/h \rfloor$ . This implies that both  $L_1$  and  $L_2$  can be set to be 1 in the case of  $M = N = h$ .

**Step 2:** For  $l = 1, 2, \dots, L$ , set  $g_l(u, v) = G_{y+v}(x+u)$  ( $u = 1, 2, \dots, h, v = 1, 2, \dots, h$ ), where  $x = [(l-1) \bmod h] \times h, y = \lfloor (l-1)/3 \rfloor \times h$ , and then compute the power spectral density  $P_l(u, v)$  from  $g_l(u, v)$  using two-dimensional FFT (2D FFT).

**Step 3:** Calculate the average of  $P_l(u, v)$  ( $u = 1, 2, \dots, h, v = 1, 2, \dots, h$ ), which is given by

$$Y(u, v) = \frac{1}{L} \sum_{l=1}^L P_l(u, v). \quad (2)$$

**Step 4:** Calculate the average of  $Y(r, \theta(d))$  for given value of  $r$ , which can be given by

$$\bar{Y}(r) = \frac{1}{180} \sum_{d=0}^{179} Y(r, \theta(d)), \quad (3)$$

where  $(r, \theta(d))$  denotes the polar coordinates of  $(u, v)$  and  $\theta(d) = 180d/\pi$ .

**Step 5:** Find the frequency components, denoted by  $r = r_{(i)}$  ( $i = 1, 2, \dots, \tau$ ), corresponding to the local maxima of  $\bar{Y}(r)$  in Eq. (3), where  $r_{(1)} < r_{(2)} < \dots < r_{(\tau)}$  and  $3 \leq \tau < h/2$ . Let  $\mathbf{x}_l = (x_1, x_2, \dots, x_\tau) = (r_{(1)}, r_{(2)}, \dots, r_{(\tau)})$ , which will be used to extract chord progressions in Section 4.

**Step 6:** Compute the scaling exponent  $\alpha_2$  as the slope of the plot  $\{\log(r)$  versus  $\log(\bar{Y}(r))\}$ .

### 2.3 Modification of Method 2 (Method 3)

We have provided the method for classifying pictures by combining the two methods using 1D and 2D FFTs[9]. This section develops the method using only 2D FFT to simplify the procedure based on the following observations:

Figure 1 shows examples of pictures, which are classified into three groups in accordance with the human subjective evaluation. Groups A, B and C consist of the pictures that categorized into “Random”, “ $1/f$ ” and “Red”, respectively. Figure 1(A1) and (B1) are

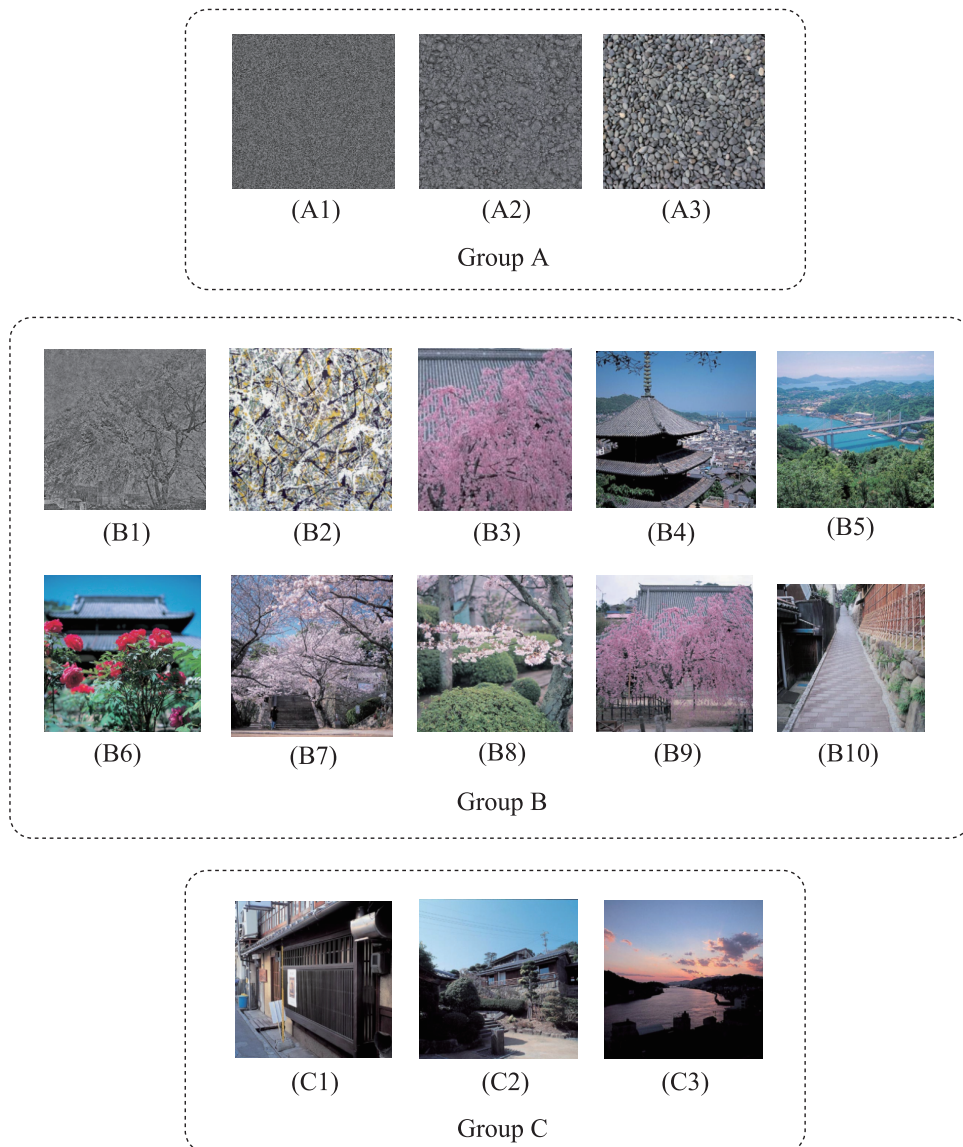


Figure 1: Examples of pictures grouped in accordance with the subjective evaluation

Table 1: Result of preliminary experiment

Picture	Subjective evaluation	Method 1 (1D)	Method 2 (2D)
(A1) - (A3)	Random	Random	Random
(B1) - (B3)	$1/f$	Random	$1/f$
(B4) - (B5)	$1/f$	Random	Red
(B6) - (B10)	$1/f$	$1/f$	Red
(C1) - (C3)	Red	Red	Red

generated by a computer as test image data so that they can include features of a random and  $1/f$  components, respectively.

Table 1 summarizes the results of a preliminary experiment using the pictures in Figure 1. Table 1 signifies that the results obtained by Method 2 agree with those obtained by the subjective evaluation for Pictures from (A1) to (A3), from (B1) to (B3) and from (C1) to (C3) in Figure 1. For Pictures from (B4) to (B10) in Figure 1, in contrast, Method 2 classifies them into “Red” while Method 1 correctly categorizes them into “ $1/f$ ”. Method 2 tends to categorize the pictures with sky or sea in Group B into “Red” since it sensitively reacts on flat-regions (sky or sea) in the pictures. The main cause of this discrepancy is the use of the average of power spectral density  $\bar{Y}(r)$  in Equation (3) to obtain the value of  $\alpha_2$ . Table 1 also shows that Method 1 incorrectly classifies Pictures (B4) and (B5) into “Random”. Method 1 is not suited to detect a self-similarity on the surface of the picture in that it analyzes the feature of the image concentrating on the brightness oscillations in the horizontal direction.

In order to modify Method 2, this subsection evaluates a content ratio of  $1/f$  components in  $Y(u, v)$  through obtaining each scaling exponent corresponding to  $\theta$ , denoted by  $\alpha(\theta) (\leq 0)$ .

We here introduce the following index:

$$C = \frac{1}{180} \sum_{d=0}^{179} |1 + \alpha(\theta(d))|, \tag{4}$$

which is considered as “Mean Absolute Deviation” when the mean value is assumed to be  $-1$ . Equation (4) signifies that  $C$  decreases as the content ratio of  $1/f$  components in  $Y(u, v)$  increases, in contrast,  $C$  increases when its ratio becomes small.

We selected 102 pictures at random to obtain the values of  $\alpha_1$  and  $C$ . The correlation coefficient  $r$  for  $\alpha_1$  against  $C$  was 0.83, which was greater than that for  $\alpha_1$  against  $\alpha_2$  ( $r = 0.77$ ). The equation of regression line obtained from these data was:

$$\hat{\alpha} = 0.77C - 2.05. \tag{5}$$

We suppose that  $\alpha_1$  can be explained by  $C$ . On the basis of above observations, the following procedure can be conducted:

**Step 1:** Calculate  $\alpha_2$  by Method 2. If  $\alpha_2 > -1.50$ , then go to Step 2, otherwise classify the picture in accordance with the values of  $\alpha_2$ :

- Classify the picture into “Random” if  $\alpha_2 \geq -0.87$ .
- Classify the picture into “ $1/f$ ” if  $\alpha_2 < -0.87$ .

**Step 2:** Calculate  $\hat{\alpha}$  by Equation (5). Classify the picture into “ $1/f$ ” if  $\hat{\alpha} \geq -1.50$ , otherwise classify the picture into “Red”.

### 3 Results of Classifying Images

This section analyzes 189 photographs published in the book edited by Onomichi city[13] excluding the pictures using in Subsection 2.3.

By subjective evaluation, 158 pictures in this book were classified into “ $1/f$ ”, remaining 31 pictures were categorized into “Red”, and there are no pictures that were classified

Table 2: Results

Category	Subjective evaluation	Method 3	Rate of correct answers
1/ <i>f</i>	158	157	96.2%
Red	31	32	83.9%

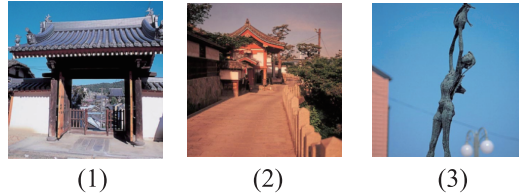


Figure 2: Example of pictures incorrectly classified by Method 3

Table 3: harmonic functions in C major

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

into “Random” as shown in Table 2. Table 2 also summarizes the number of the pictures classified into each category by Method 3 and the rate of correct answers. Table 2 shows that 96.2% and 83.9% of the pictures could correctly classified into “1/*f*” and “Red”, respectively (94.2% on the whole). This signifies that Method 3 can effectively categorize the pictures of tourist destinations in a local city.

Figure 2 shows an example of pictures that were incorrectly classified by Method 3. These pictures in Figure 2 were grouped into “1/*f*” by the subjective evaluation, in contrast, by Method 3, they were categorized into “Red”. These pictures consist of the scenery that artificial objects and sky harmonize well. Method 3 reacts on flat-regions (sky or sea) in the picture even if the picture has well composition.

## 4 Music Extraction from Images

### 4.1 Procedure for extracting chords

We have already proposed the procedure for extracting chords using  $\mathbf{x}_l$ [9], defined in Section 2. Let  $\mathbf{y}_l$  ( $l = 1, 2, \dots, L$ ) express the musical chord in relation to the square  $l$  obtained by this procedure.

### 4.2 Method for finding chord progressions

This subsection proposes a method for finding chord progressions using the chords obtained in Section 4.1.

We consider the following conditions for simplicity:

- (i) The chord progressions are generated in C major.
- (ii) 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 a natural tone. Let  $L' (\leq 0)$  denote the number of the chords selected here.
- (iii) The chords are arranged so that the chord progression can include cadences (T-S-T, T-D-T and T-S-D-T). Table 3 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 4.1, the chord progression of C-F-G-C (T-S-D-T) is then obtained. We here introduce some more additional notation to define an evaluating function below.

**q:**  $\mathbf{q} = (q_1, q_2, \dots, q_{L'})$ , where  $q_i \in \{1, 2, \dots, L'\}$  expresses a number corresponding to each chord selected from  $y_l$  ( $l = 1, 2, \dots, L'$ ) under Condition (ii).

- $p_1(\mathbf{q})$ : the number of T-D-T in a chord progression.
- $p_2(\mathbf{q})$ : the number of T-S-D-T in a chord progression.
- $p_3(\mathbf{q})$ : the number of T-S-T in a chord progression.
- $p_4(\mathbf{q})$ : the number of strong cadences in a chord progression.

The evaluating function is defined by

$$P(\mathbf{q}) = \sum_{i=1}^4 w_i p_i(\mathbf{q}), \tag{6}$$

where  $w_i$  ( $0 \leq w_i \leq 1$ ) is a weight coefficient for  $p_i(\mathbf{q})$ .

We consider the problem for sorting the  $L'$  chords under Condition (iii). The problem can be formulated as follows:

$$\left. \begin{array}{l} \text{maximize } P(\mathbf{q}) \\ \text{subject to } q_i \in \{1, 2, \dots, L'\}, \\ q_i \neq q_j \ (i, j = 1, 2, \dots, L', i \neq j). \end{array} \right\} \tag{7}$$

Problem (7) is NP-hard since finding an optimal solution to (7) strictly needs to solve this at  $L'!$  times. The solution can be obtained by applying meta-heuristics such as the tabu search algorithm[12].

## 5 Examples of method for extracting chord progressions

This section presents simple examples for the extraction of chord progressions from some pictures to illustrate the proposed method in this study. We consider the case where each picture is divided into 16 squares, which indicates that less than or equal to 16 chords are extracted from a picture as mentioned in the previous work[9]. When the size of  $\tau$ , denoted in Step 5) of Subsection 2.2, is too small to extract a chord, the procedure described in Subsection 4.1 will output “none”, which indicates that a chord cannot be extracted from the square in the picture.

(a) Example 1

(b) Example 2

(c) Example 3

Figure 3: Example of generated chord progressions

By solving Problem (7) applying the tabu search algorithm [12], chord progressions as shown in Figure 3(a), (b) and (c) are generated from Figure 1(A3), (B3) and (C3), respectively. As mentioned in Subsection 2.3, Figure 1(A3), (B3) and (C3) were respectively categorized into “white noise”, “ $1/f$  noise” and “red noise”.

The chord progression in Figure 3(a) consists of many chords containing dissonant intervals such as minor 6th chord. Figure 1(A3), corresponding to this progression, includes the structure of “white noise”, having (almost) equal intensity at difference frequencies. In this case, the difference between two adjacent elements of  $\mathbf{x}$ , denoted in Step 5 of Subsection 2.2, becomes small, which leads to a narrowing of two adjacent sounds in the generated chord.

In contrast, the chord progression in Figure 3(b) and (c) consist of the chords in which the intervals between two adjacent sounds are adequate, reflected by the behavior of the power spectral densities obtained from Figure 1(B3) and (C3).

Figure 3 also shows that the length of the chord progression decreases in order from (a) to (c). This reason why the number of peaks of the power spectral densities used to extract a chord tends to decrease in order of Figure 1(A3), (B3) and (C3), having the structures of “white noise”, “ $1/f$  noise” and “red noise”, respectively. The scaling exponent  $\alpha_2$ , defined in Step 6 of Subsection 2.2, decreases in order of “white noise”, “ $1/f$  noise” and “red noise” as mentioned in Section 2, which signifies that the range of frequency to extract a chord becomes narrower in order of Figure 1(A3), (B3) and (C3).



## 6 Conclusion

This study has developed the following two methods applying the 2D FFT (fast Fourier transform) algorithms: (1) a method for finding pleasant photographs of local tourist destinations based on the results of a preliminary experiment, and (2) a method for generating chord progressions from these photographs considering the two kinds of “intervals”: (a) the difference between the frequency components and (b) the difference between pitches in music theory.

For a preliminary experiment, 300 pictures of test image data and pictures of tourist destinations were classified into three groups, Group A, B and C, corresponding to “Random”, “ $1/f$ ” and “Red”, respectively, by the subjective evaluation based on interviews with students, professors of arts and civils. We have confirmed that the method applying the 2D FFT (Method 2) can correctly classifies the above pictures except for the pictures in Group B in which the sky, the ocean or the lake appear. The contrast of these objects are relatively low and flat, and therefore Method 2 incorrectly classifies them into “Red” in that it is affected by the monotone areas in the picture. In contrast, the method applying 1D FFT (Method 1) can correctly categorize them, while it tends to incorrectly classify the pictures with a self-similarity on the surface into “Random” since it focuses on the brightness oscillations in the horizontal direction . Furthermore, we clarifies a relationship between the scaling exponent associated with Method 1 and a content ratio of  $1/f$  components in the power spectrum densities obtained from Method 2. On the basis of the above observations, we provided a modified method applying only 2D FFT to simplify the procedure. The results of experiments indicate that the method proposed here can identify the pictures of the tourist destinations as “ $1/f$  or “Red” according to our intuition.

This study also proposed a method for generating a chord progression from a picture. We could extract chords which “embed” or “hidden” in the picture considering the inversion and the interval in musical theory. Some examples were provided to illustrate the proposed method. In this study, the chords were extracted by applying 2D FFT, but by using the wavelet analysis, a wide variety of chord progressions can be generated from a picture. Taking account of such factors is an interesting extension.

## References

- [1] R. Botha and C. Knight, *The Prehistory of Language*. Oxford University Press, USA, 2009.
- [2] R. F. Voss and J. Clarke, ““ $1/f$  noise” in music: Music from  $1/f$  noise,” *J. Acoust. Soc. Am.*, vol. 63, no. 1, 1978, pp. 258–263.
- [3] 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.
- [4] 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, 2007, pp. 570–584.
- [5] W. Ro and Y. Kwon, “ $1/f$  noise analysis of songs in various genre of music,” *Chaos Solitons & Fractals*, vol. 42, 2009, pp. 2305–2311.

- [6] E. Rodriguez *et al.*, “ $1/f$  noise structures in Pollocks’ drip paintings,” *Physica A*, vol. 387, 2008, pp. 281–295.
- [7] T. Musha, *Yuragi no Sekai [The World of Fluctuations]*, Kodansha, 1980.
- [8] H. Kawakatsu, “Methods for evaluating pictures and extracting music by 2D DFA and 2D FFT,” *Procedia Computer Science*, vol. 60, 2015, pp. 834–840.
- [9] H. Kawakatsu, “Methods for Classifying Pictures of Tourist Destinations and for Transforming Frequency Components into Musical Chords,” Proc. of 5th International Congress on Advanced Applied Informatics (AAI 2016), 2016, pp. 807–812.
- [10] E. Rodriguez, J. C. Echeverria, and J. Alvarez-Ramirez, “ $1/f^\alpha$  fractal noise generation from Grünwald-Letnikov formula,” *Chaos, Solitons & Fractals*, vol. 39, no. 2, 2009, pp. 882–888.
- [11] H. Kawakatsu, “Fluctuation analysis for photographs of tourist spots and music extraction from photographs,” Lecture Notes in Engineering and Computer Science: Proc. of The World Congress on Engineering 2014 (WCE2014), vol. 1, 2014, pp. 558–561.
- [12] H. Katagiri, T. Hayashida, I. Nishizaki, and Q. Guo, “A hybrid algorithm based on tabu search and ant colony optimization for k-minimum spanning tree problems,” *Expert Systems with Applications*, vol. 39, no. 5, 2012, pp. 5681–5686.
- [13] Onomichi city, *Onomichi Sansaku 204 Sen [Selected 204 exquisite spots in Onomichi]*. Toppan Printing, 2004.