Methods for Classifying Pictures and Generating Music by 2D DFA and 1D FFT

Hidefumi Kawakatsu*

***Abstract*—This study proposes the following two methods applying two-dimensional DFA (detrended fluctuation analysis) and one-dimensional FFT (fast Fourier transform) algorithm: (1) a method for finding pleasant photographs of local tourist spots, and (2) a method for creating music from these photographs. We define "pleasant photograph" as the photograph containing $1/f$ noise components since it has been suggested that the $1/f$-noise structure in visual art as well as in music can stimulate the perception of pleasant. We analyze 198 photographs published in the book edited by Onomichi city to find the pleasant photographs. The method for the music extraction from the picture is developed taking into account that the sequence of the brightness in the horizontal direction on each row can be decomposed into periodic waves, and then, by regarding each wave obtained as sound waves, successive chords can be created from the image. A method to arrange the music by reflecting features of the image is also provided.**

***Index Terms*—detrended fluctuation analysis, fast Fourier transform, $1/f$ noise, pleasant photograph, music extraction**

I. INTRODUCTION

Local cities in which there are many resources for tourism attempt to increase tourist arrivals. This study provides a way helping to recommend tourist attractions. 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[1], [2], [3], [4]. They indicated that pleasant music for humans displays a behavior similar to $1/f$ noise. For visual arts, E. 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 than the two-dimensional FFT (fast Fourier transform), but its computation time becomes greater than that of the one-dimensional case. Musha[6] has described the method which evaluates paintings by using one-dimensional Fourier transform as follows: (1) computing the power spectral densities of brightness in the horizontal direction in each row, and then (2) obtaining the average values for the spectral densities in each column. In our previous work, we propose a method for finding pleasant photographs of tourist spots applying the one-dimensional FFT[7] for saving the computational time. In this study, we propose the method applying the two-dimensional DFA algorithm to compare with the results which were obtained from our previous method[8].

Manuscript received December 11, 2014; revised January 27, 2015.
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.617); fax: +81-848-22-5460; e-mail: kawakatsu@onomichi-u.ac.jp).

We have also developed the method for extracting music by using the one-dimensional FFT algorithm[8]. This study modifies our method in order to generate various sort of chord. The melodies are generated considering the fact that the brightness of the picture can be resolved into periodic waves, and then, by regarding each wave obtained from the picture as sound waves, successive chords can be created from the image.

Our methods may be available to recommend a sightseeing route which connects the pleasant spots for tourists, and to create the music from photographs of these spots.

II. 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:

$$P(f) \sim 1/f^{\alpha_f}, \tag{1}$$

where α_f is called the scaling (spectral) exponent. The scaling exponent α_f is computed as the slope of the plot $\{\log(f) \text{ versus } \log(P(f))\}$. The sequence, denoted by $u(i)$, can be classified as follows: (a) If $\alpha_f = 0$, the sequence $u(i)$ is non-correlated (i.e. white noise); (b) If $\alpha_f = -1$, the correlation of the sequence is the same of $1/f$ noise; and (c) If $\alpha_f = -2$, the sequence behaves like Brownian motion (i.e. red noise)[1], [9].

Self-similarity of the signal can also be detected by DFA (detrended fluctuation analysis)[2], [3], [5]. The scaling exponent α_d can be given by applying the DFA method, which will be explained in Section II-B. In this case, the profile, defined by $y(i) = \sum_{k=1}^{i}[u(k) - \bar{u}]$, displays properties of the white noise with $\alpha_d = 0.5$. The values $\alpha_d = 1.0$ and $\alpha_d = 1.5$ respectively indicate $1/f$ noise and Brownian noise.

We define "pleasant photograph" as the photograph containing $1/f$ components and propose a procedure for finding the pleasant photographs using the (one-dimensional) fast Fourier transform (FFT) algorithm and the two-dimensional detrended fluctuation analysis (DFA). 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 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 ($k = 2, 3, \cdots$) when we use the FFT algorithm. 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.

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

A. Method using 1D FFT

We have already proposed the method for classifying the pictures applying the one-dimensional FFT algorithm[8]. We define a_f as the scaling (spectral) exponent obtained by this method. In this case, the pictures can be classified into one of the following three cases: (a) If $\alpha \geq -0.75$, the image is classified into "Random" (or "White"): (b) If $-1.25 \leq \alpha < -0.75$, the image is classified into "$1/f$ noise" (or "Pink"): (c) If $\alpha < -1.25$, the image is classified into "Brownian" (or "Red").

B. Method using 2D DFA

This section proposes the method for classifying the pictures using the two-dimensional DFA algorithm[5]. In order to simplify the discussion, we confine ourselves to the case where $q = 2$.

1) Calculate the profile, expressed by $x_{i,j}$, which is given by
$$x_{i,j} = \sum_{n=1}^{i} \sum_{m=1}^{j} \left[G_i(j) - \bar{g}\right], \tag{2}$$
where $\bar{g}$ represents the mean value of $G_n(m)$.
2) Divide $x_{i,j}$ into small regions of size $s \times s$, where s is set to be $s_{\min} \simeq 5 \leq s \leq s_{\max} \simeq \min\{M, N\}/4$.
3) Compute an interpolating curve $\gamma_{i,j}(l,s) = a_l i + b_l j + c_l$ of $x_{i,j}$ in the lth small square region of size $s \times s$, which can be given by using a multiple regression procedure.
4) Calculate the "variance" in the lth small square region for $s = s_{\min}, s_{\min}+1, \cdots, s_{\max}$, which is given by
$$F_{i,j}^2(l,s) = \frac{1}{s^2} \sum_{n=i}^{i+s} \sum_{m=j}^{j+s} z_{i,j}^2(l,s), \tag{3}$$
where $z_{i,j}(l,s)$ is expressed by
$$z_{i,j}(l,s) = x_{i,j} - \gamma_{i,j}(l,s). \tag{4}$$
5) Compute the root mean square, denoted by $F(s)$, of $F_{i,j}^2(l,s)$, which is given by
$$F(s) = \left[\frac{1}{L_s} \sum_{l=1}^{L_s} F_{i,j}^2(l,s)\right]^{1/2}, \tag{5}$$
where L_s denotes the number of the small square regions of size $s \times s$.
6) If $x_{i,j}$ has a long-term correlation property, a power-low behavior for the fluctuation function $F(s)$ is observed as follows:
$$F(s) \sim \alpha_d^s, \tag{6}$$
where α_d is called the (two-dimensional) scaling exponent, a self-affinity parameter representing the long-range power-low correlation properties of the surface. This scaling exponent α_d is computed as the slope of the plot $\{\log(s) \text{ versus } \log(F(s))\}$.

In this study, we classify the picture into one of the following three cases: (a) If $\alpha_d \leq 0.75$, the image is classified into "Random" (or "White"): (b) If $0.75 < \alpha_d \leq 1.25$, the image is classified into "$1/f$ noise" (or "Pink"): (c) If $\alpha_d > 1.25$, the image is classified into "Brownian" (or "Red").

III. Music Extraction from Images

We have proposed a simple method for extracting music from the image using the one-dimensional FFT algorithm[8]. This section modifies our method in order to obtain various sort of chord.

The image is transformed into the gray scale image $G_n(m)$, and the size of the image is normalized to M (horizontal) $\times N$ (vertical) pixels as mentioned in Section II.

The horizontal axis of the image can be regarded as the time axis by using the following equation.
$$t = m/S_f, \; m = 1, 2, \cdots, M \tag{7}$$
where t represents the time corresponding to m and S_f expresses a sampling frequency. We assign an integer to each musical note of the scale such that middle C (C4) is assigned 37, the note just above (C#4) is 38 and the musical rest is 0. In this case, the frequency F_j (Hz) of each sound can be given by
$$F_j = \begin{cases} 440 \times 2^{\frac{j-46}{12}}, & \text{if } j = 1, 2, \cdots, 96, \\ 0, & \text{if } j = 0. \end{cases} \tag{8}$$

Let H $(\leq b - a)$ be the maximum value of the number of sounds (notes) per chord, then the following procedure can be used to extract the music from the gray scale image:

1) $n = 1$.
2) Set $u(m) = G_n(m)$ $(m = 1, 2, \cdots, M)$, and pass $u(m)$ through a bandpass filter in the range from a Hz to b Hz, where $a = 0.015M$ and $b = 0.35M$. The values of a and b are determined considering the Nyquist frequency (folding frequency) and the distortion of the spectrum.
3) Compute the spectral density $s_n(f)$ of the sequence $u(m)$ using the FFT algorithm. If $n < N$, increment n by one and go to Step 2), otherwise go to Step 4).
4) Set a suitable value for S_f.
5) $l = 1$.
6) Divide the image into L $(\leq N)$ regions from the first row of the image (The size of each region becomes $\lfloor N/L \rfloor$ (vertical) $\times M$ (horizontal)).
7) Compute the average of the power spectral density $\bar{P}_l(m)$ in the lth $(l = 1, 2, \cdots, L)$ region for $m = 1, 2, \cdots, M$, which is given by
$$\bar{P}_l(m) = \frac{1}{w} \sum_{k=(l-1)w+1}^{l \cdot w} s_k^2(m). \tag{9}$$
8) Compute the interpolating curve, denoted by $\hat{P}_l(m)$, of $\bar{P}_l(m)$ applying the regression analysis. Define the set A as follows:
$$A = \{m \mid \bar{P}_l(m) \geq \hat{P}_l(m)\}. \tag{10}$$
9) $i = 1$.
10) Compute the frequency f_i corresponding to m'_i $(m'_i \in A$, $m'_i \neq m'_j$ for $i \neq j)$ and the absolute value d_j of the difference between f_i and F_j, where $f_i = m'_i \cdot (S_f/M)$ and $d_j = |f_i - F_j|$ $(j = 1, 2, \cdots, 96)$.
11) Find F_k such that $d_k = \min d_j$ and let $q_{i,l} = k$ be a note in the chord $\boldsymbol{Q}(l)$ which is generated from lth region, where $\boldsymbol{Q}(l)$ is expressed by Eq. (11) in the next step.

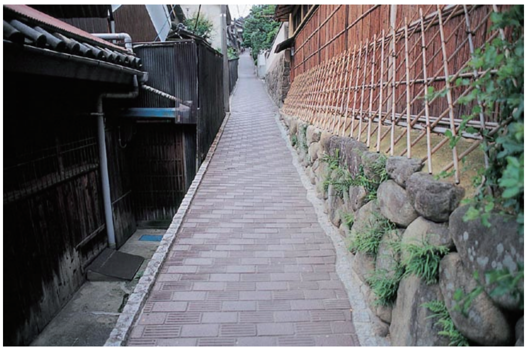

Fig. 1. Example of results(a)

12) If $i < H$, increment i by one and go to Step 10), otherwise set

$$\boldsymbol{Q}(l) = (q_{1,l}, q_{2,l}, \cdots, q_{H,l})^T \tag{11}$$

and go to Step 13).
13) If $l < L$, increment l by one and go to Step 7), otherwise go to Step 14).
14) Arrange $\boldsymbol{Q}(l)$ in order from $\boldsymbol{Q}(1)$ to $\boldsymbol{Q}(L)$ and transform them into the musical score.

We refer to the music generated from above procedure as an "original music". A method for arrangement the original music by reflecting features of the image will be discussed in Section V.

IV. Results of classifying images

In our previous study[8], we have analyzed the 198 photographs published in the book edited by Onomichi city[10] by means of the method applying the one-dimensional FFT algorithm (*Method 1*). This section analyzes the same pictures using the method proposed in Subsection II-B (*Method 2*). We show the features of Method 1 and 2, by focusing on the pictures that different results are obtained from the two methods. Results obtained from Method 1 and 2 are summarized in Table I in the case where we analyze pictures in Figures from 1 to 4.

Table I shows that the pictures in Figs. 1 and 3 are classified into "$1/f$ " and "Red " using Method 1 and 2, respectively. As mentioned in our previous work[8], the behavior of the brightness in the horizontal direction on the pictures in Figs. 1 and 3 displays that of similar to $1/f$-noise. The Method 2, in contrast, classifies them into the category of "Red". Pictures with low and flat contrast and with large dark area are tend to be classified into this category as mentioned in our previous study[8]. Figures 2 and 4 respectively display cropped images from Figs. 1 and 2. They are trimmed so as not to include dark and flat area. Table I also indicates as follows: (i) the cropped image in Fig 2 is classified into "Red " and "$1/f$ " from Method 1 and 2, respectively. (ii) the cropped image in Fig 4 is classified into "$1/f$ " (close to "White ") and "$1/f$ " from Method 1 and 2, respectively. This signifies that Method 2 is suitable for detecting a self-similarity on the surface of images.

V. Arrangement of "original music"

This section develops a method of arrangement of the "original music" which is created by the procedure proposed

TABLE I
Results

Picture	a_f	a_d
Fig. 1	-1.263 ($1/f$)	1.676 (Red)
Fig. 2	-1.675 (Red)	1.238 ($1/f$)
Fig. 3	-0.815 ($1/f$)	1.776 (Red)
Fig. 4	-0.775 ($1/f$)	1.161 ($1/f$)

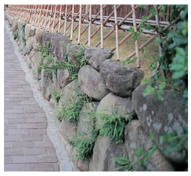

Fig. 2. Example of results(b)

Fig. 3. Example of results(c)

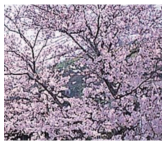

Fig. 4. Example of results(d)

in Section III. We consider the case where, as the average value of the brightness in the region l ($l = 1, 2, \cdots, L$) increases, the number of sounds (notes) in the chord $\boldsymbol{Q}(l)$ decreases, tempo increases and the ratio of containing dissonance per chord decreases. By this way, a part of the music extracted from the bright area in the picture can be played lightly, while it is played heavily corresponding to the brightness of the dark area.

We also consider the case where some sounds are interpolated into two successive chords $\boldsymbol{Q}(l)$ and $\boldsymbol{Q}(l+1)$ ($l = 1, 2, \cdots, L-1$) reflecting the scaling (spectral) exponent in the region l.

We here introduce some notation with respect to the region l ($l = 1, 2, \cdots, L$) in the picture for simplicity as listed below:

$\bar{g}_l$: the average value of the brightness in the region l.
α_l: the scaling (spectral) exponent.
$\tau(l)$: the tempo of the part corresponding to the region l.
$\theta(l)$: the number of tones sounded simultaneously.
$\lambda(l)$: the ratio of containing dissonance.

Let $q_{H,l}$ correspond to a melody part and the other sounds in $\boldsymbol{Q}(l)$ without $q_{H,l}$ be an accompaniment part.

In this case, the procedure for the arrangement of the original music can be proposed as follows:

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

1) $l = 1$.
2) Compute the average value $\bar{g}_l$ of the brightness and the scaling exponent α_l in the region l.
3) Set the values for $\tau(l)$, $\theta(l)$ and $\lambda(l)$ such that these values are proportional to $\bar{g}_l$, where $\tau(l)$, $\theta(l)$ and $\lambda(l)$ take the values on the interval $[40, 255]$, $[0, 0.1]$ and $[1, H]$, respectively.
4) Adjust the values for the components of $\boldsymbol{Q}(l)$ without $q_{H,l}$ such that the ratio of containing dissonance in $\boldsymbol{Q}(l)$ is less than or equal to $\lambda(l)$.
5) Set $\boldsymbol{S}((l-1)(B+1)+1) = \boldsymbol{Q}(l)$, where B denotes the number of interpolating sounds.
6) If $l = 1$, go to Step 9), otherwise go to to Step 7).
7) For the melody part, set the value of $s_{H,(l-2)(B+1)+j}$ $(j = 2, \cdots, B-1)$ corresponding to the value for α_l as follows:
   a) $\alpha_l \geq -0.75$.
      Set
      $$s_{H,(l-2)(B+1)+j} = r^a_{j-1}, \tag{12}$$
      where r^a_k $(k = 1, 2, \cdots, B-2)$ is a positive integer, which is given as the uniform random number on the interval $[q_{H,l-1}, q_{Hl}]$.
   b) $-1.25 \leq \alpha_l < -0.75$.
      Set
      $$s_{H,(l-2)(B+1)+j} = r^b_{j-1}, \tag{13}$$
      where r^b_k is a positive integer, which is given as the $1/f$ random number with mean $(q_{H,l-1} + q_{Hl})/s$.
   c) $\alpha_l < -1.25$.
      Set
      $$s_{H,(l-2)(B+1)+j} = \begin{cases} s_{H,(l-2)(B+1)+j-1} - 1, & p_{j-1} < 0.5 \\ s_{H,(l-2)(B+1)+j+1} + 1, & p_{j-1} \geq 0.5 \end{cases} \tag{14}$$
      where p_k expresses the uniform random variable on the interval $(0, 1)$.
8) For the accompaniment part, set $s_{ij} = qil$ $(i = 1, 2, \cdots, H-1,\ j = (l-1)(B+1)+2, (l-1)(B+1)+3, \cdots, (B+1)l)$.
9) If $l < L$, increment l by one and go to Step 2), otherwise go to Step 10).
10) Arrange $\boldsymbol{S}(l)$ in order from $\boldsymbol{S}(1)$ to $\boldsymbol{S}(L)$ and transform them into the musical score.

VI. Conclusion

This study has developed the following two methods applying the two-dimensional DFA (detrended fluctuation analysis) and the one-dimensional FFT (fast Fourier transform) algorithm: (1) a method for finding pleasant photographs of local tourist spots using the two-dimensional DFA, and (2) a method for creating music from these photographs using one-dimensional FFT. We have defined "pleasant photograph" as the photograph containing $1/f$ noise components since the recent studies (such as E. Rodriguez et al. [5]) had suggested that paintings which contain $1/f$-noise structures stimulate the perception of pleasant.

In our previous study[8], we analyzed the 198 photographs published in the book edited by Onomichi city[10] by means of the method applying the one-dimensional FFT algorithm (*Method 1*). This study have analyzed the same pictures using the method proposed in Subsection II-B (*Method 2*) to compare the features of Method 1 and 2. In this study, we also have modified our method for extracting music[8] so as to generating various kind of chord.

When we had shown the results obtained from the analysis using the 198 pictures at a seminar for the general public held in Onomichi, and asked for the participants' opinions, they mostly agreed with our observation. This indicates that our method can efficiently be applied to classify the pictures of tourist areas where scenery with old houses, shrines and temples. In contrast, the comment on music created from these pictures was that music should be created enough to evoke the scene in the pictures. The pictures used in this study are transformed into gray scale images. In many cases, the pictures of the tourist spots where countryside scenery are preserved consist of mountains, plants, coasts, temples and so on. The color of mountains, coasts and temples are usually green, blue and brown, respectively. Various melodies can be created by changing of musical key depending on the color of the objects in 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] T. Musha, *Yuragi no Sekai [The World of Fluctuations]*. Kodansha, 1980.
[7] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, 2000.
[8] H. Kawakatsu, "Fluctuation analysis for photographs of tourist spots and music extraction from photographs," *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2014, WCE 2014, 2-4 July, 2014, London, U.K.*, vol. 1, pp. 558–561, 2014.
[9] 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, pp. 882–888, 2009.
[10] Onomichi city, *Onomichi Sansaku 204 Sen [Selected 204 exquisite spots in Onomichi]*. Toppan Printing, 2004.