

The Color-Music Relationship Direct Modeling using Machine Learning

Jong Yeol Lee
Department of Computer Software
Engineering
Kumoh National Institute of
Technology
Gumi, South Korea
soyeum@kumoh.ac.kr

Chang Bae Moon
ICT-Convergence Research Center
Kumoh National Institute of
Technology
Gumi, South Korea
cb.moon@kumoh.ac.kr

Byeong Man Kim
Department of Computer Software
Engineering
Kumoh National Institute of
Technology
Gumi, South Korea
bmkim@kumoh.ac.kr

Abstract— As a method to improve the effectiveness of music retrieval, this paper suggests a method that can exclude subjective judgment in music retrieval by directly modeling the color-music relationship and can assure perpetuity by automating color discernment on new music. To this end, a method to define representative color of each music from colors that people reminisce after hearing music is presented and color-music model that classifies feature features of music by using machine learning is also presented. The experiment shows performance improvement by 4% compared with VGG-19.

Keywords—*Color-Music Relationship; Music Classification; Music Retrieval; Convolutional Neural Network; Deep Learning*

I. INTRODUCTION

Recently, the cases that increase learning effect or sales by using variation in human emotion through music in the areas like education, marketing, psychological treatment tends to increase. For example, letting babies hear music in the area of education, positive effect is shown in the development of cognitive ability, intellect and brain[1]. In marketing area, providing music suitable to the mood of relevant shop like cafe or apparel shop causes positive effect to increase sale[2]. The significant factor is to select music proper to current situation.

Music selection methods can be divided into 2 groups, music recommendation by expert and general music retrieval. As music recommendation by expert is costly, general music retrieval can be realistic in consideration of cost. However, general music retrieval may decrease effectiveness due to the subjective judgment of a retriever, and perpetuity cannot be assured due to the release of new music.

This paper suggests a method to improve effectiveness of music retrieval and to assure perpetuity caused by new music. Namely, as a method to improve effectiveness of music retrieval, color-music relationship is modelled to exclude subjective judgment in music retrieval and enhance effectiveness. The method to assure perpetuity by automatically discerning color for new music is presented. To this end, this paper presents color-music relationship model that classifies features of music by using machine learning

after defining representative color of each music using colors that people reminisce after hearing music.

II. RELATED STUDIES

Palmer, et al let American and Mexican applicants hear classical music of Bach, Mozart, Brahms to select color. Hearing 1 piece of music, they are told to select color they thought most and least relevant, 5 colors respectively out of 38 colors. Then, music was analyzed in the division of 4 kinds of mode and tempo. In general, they selected low saturation, low brightness and blue series for slow music and high saturation, high brightness and yellow series for fast music[3].

Barbiere, et al studied how to designate color for musical mood by using 2 types of mood (happy or sad and 11 basic colors. It was found that bright color was generally related to happy song and gray was generally related to sad song [4].

Bresin, et al confirmed that color and mood are mutually related by grasping color-music relationship by using 12 songs which produce mood different from 8 colors. Namely, dark color is found to be related to music in minor key and bright color to music in major key[5].

Using information on music genre as pre-step of music-color relationship, Moon Chang Bae, et al arranged the feature of mood and color distribution and grasped relationship between mood and color depending on music genre as a part of studying method to express mood by using hearing and vision. To this end, they analyzed mood distribution and color distribution depending on music genre by receiving mood of sound source and color of mood from subject through online question papers and proved that there was difference in mood and color depending on music genre[6].

This paper uses Convolution Neural Network (CNN) as color-music model. Same performance is expected in sound source due to excellent feature extraction and discerning ability for new high dimensional image. Convolution Neural Network (CNN) is deep neural network (DNN) in which layers that basically play 3 different roles are repetitiously combined. Features are extracted from high dimensional data through convolutional layer and pooling layer, and a

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2017R1D1A1B03033733, 2018R1C1B6001042).

classifying model is constructed through fully connected layer[7, 8].

III. DIRECT MODELING OF COLOR-MUSIC RELATIONSHIP

As shown in Fig. 1., the methods suggested in this paper are composed of 3 steps. In the 1st step, sound source to be used for color-music relationship modelling is secured and representative section for each sound source is selected to define representative color for each sound source. In the 2nd step, feature vector for each sound source is formed to use representative section for selected sound source as input of machine learning. In the final step, a classifier of Convolution Neural Network to learn color-music relationship is made using the color information given by many people and the low level feature vectors of music.

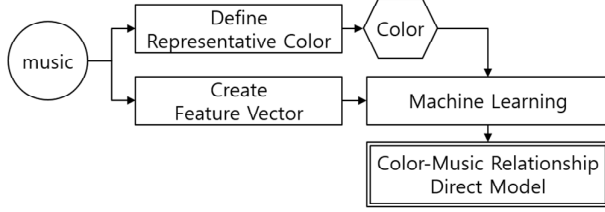


Fig. 1. Modeling Process

A. Definition of representative color for each sound source

As color data on sound source is collected from many people, respondents may select different color for same sound source. Thus, color that represents each sound source should be defined to be used as basic information. To this end, RGB color value chosen by respondents is converted to CIE Lab color value. Representative color is defined by getting average through formula (1).

$$\begin{aligned}
 L_s &= \frac{1}{n} \sum_{i=1}^n (116f(Y/Y_n) - 16) \\
 a_s &= \frac{1}{n} \sum_{i=1}^n (500[f(X/X_n) - f(Y/Y_n)]) \\
 b_s &= \frac{1}{n} \sum_{i=1}^n (200[f(Y/Y_n) - f(Z/Z_n)])
 \end{aligned} \quad (1)$$

where, L_s , a_s , b_s are the average of respondents' evaluation of sound sources and X , Y , Z is the value obtained by converting the normalized RGB to the CIE XYZ as value between 0 and 1 as described in formula (2). X_n , Y_n , Z_n is X , Y , Z value of based white color respectively.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4360747 & 0.3850649 & 0.1430804 \\ 0.2225045 & 0.7168786 & 0.0606169 \\ 0.0139322 & 0.0971045 & 0.7141733 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

The function f is calculated as shown in formula (3).

$$f(q) = \begin{cases} q^{1/3} & \text{for } q > 0.008856 \\ 7.787q + 0.137931 & \text{for } q \leq 0.008856 \end{cases} \quad (3)$$

B. Formation of feature vector for each sound source

As a process to form input vector for machine learning, the process in Fig. 2 is performed and finally forms square matrix 224×224 is generated. In general, the color recollected by respondents after hearing music is induced as stimulation of emotion for the representative section of sound source. Thus, 50-second playing section between 30-second and 80-second that usually include representative emotion is selected as section that can represent color of sound source.

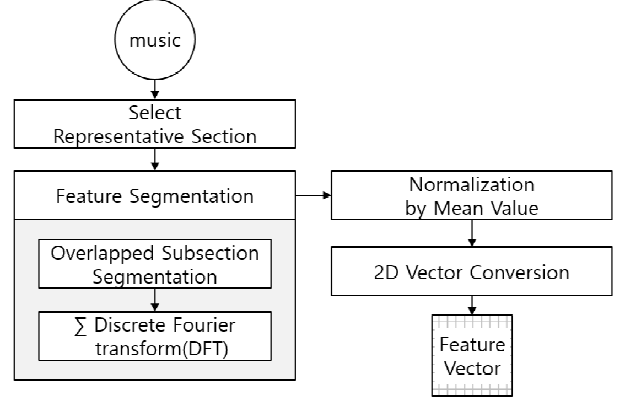


Fig. 2. Feature Vector Generation

To constitute element of feature vector, representative section is segmented into 224 sections and each section is segmented again into 224 overlapped subsections. In order to extract feature from each overlapped subsections, the discrete Fourier transform (DFT) is performed to acquire frequency feature of sound source. Through this process, the feature vector $R = \{R_1, R_2, \dots, R_{224}\}$ is obtained for the representative section of music, where each element R_i is 224-dimensional vector calculated as the accumulated sum of the overlapped subsections. Discrete Fourier transform data of sound source are normalization by their average value and standard deviation. The normalized feature vector has the 1st dimension form, which is converted to 2nd dimensional vector in the form of $(224, 224)$ in order to be used as input value of machine learning.

C. Machine learning model

CNN structure suggested in this paper is shown in Fig. 4, where, C refers to the convolution layer, M refers to the pooling layer and F refers to the fully-connected layer. Convolution layers (C1-C9) have 3x3 convolution layer (C1) for rough feature extraction. convolution layers (C2, C4, C6, C8) that have 1x1 filter summarizing feature of former layer and convolution layers (C3, C5, C7, C9) that have 3x3 filter for feature extraction. Five pooling layers(M1-M5) are laid out to reduce data size or emphasize specific data and their filters are fixed in 2x2 for max-pooling. Their stride are fixed to 1 for convolution layer and to 2 for pooling layer. Output value below 0.5 is dropout in the fully connection layer to prevent overfitting. The activation function ReLU is used in all layers except the output layer and the Softmax function is used in the output layer to determine one out of 16 colors. Each layer in

Fig. 4 is processed from left to right and from up to bottom in the order of number.

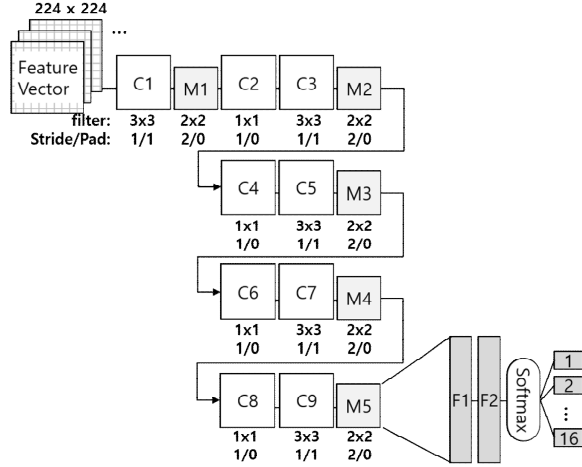


Fig. 3. The proposed CNN structure

IV. EXPERIMENTS

In order to analyze performance of suggested color-music relationship model, sound data and color data for each music piece is collected. For sound data, 50 pieces for 5 genres are prepared to be randomly provided to evaluator 10 by 10. The evaluator is told to redundantly select 3 colors among 16 colors. Respondents are divided into common people group targeting people in neighborhood and applicants group targeting common students in order to collect materials by online method. The number of people participating in investigation is 162 and 4762 color data was secured including the case which was suspended without hearing all provided 10 pieces, which means that 19 color data was secured for 1 sound source in average. Out of sound source data, 80%(200 pieces) were used for learning of suggested model and 20%(50 pieces) for verification.

Comparing CNN model suggested in this paper with 2 models of VGG, it showed better performance than VGG-19 by 4% and VGG-16 by 14% as shown in Table 1.

TABLE I. PERFORMANCE COMPARISON RESULT

| Genre | Model | VGG-19 | VGG-16 | Proposed Model |
|--------------|-------|--------|--------|----------------|
| Domestic pop | | 78% | 70% | 84% |
| Foreign pop | | 76% | 64% | 82% |
| Classic | | 86% | 76% | 90% |
| Hip hop | | 80% | 72% | 82% |
| Dance | | 84% | 74% | 88% |
| Average | | 81.2% | 71.2% | 85.2% |

The number of color for each genre of sound source selected by respondents is shown in Table 2. Where, the standard deviation is quite large except Foreign pop, which implies some color is preferred for certain genre and music

retrieval is possible by using the fact. Further, music retrieval can be more efficient if utilizing additional data such as ages, gender, occupation, the preferred genre of respondents.

TABLE II. COLORS CHOSEN BY GENRE (ALLOW DUPLICATE)

| Genre | Domestic Pop | Foreign Pop | Classic | Hip hop | Dance |
|---------------|--------------|-------------|---------|---------|-------|
| Red | 176 | 91 | 62 | 137 | 319 |
| Yellow Red | 34 | 43 | 11 | 92 | 52 |
| Gold | 27 | 54 | 5 | 21 | 31 |
| Green | 123 | 44 | 218 | 78 | 91 |
| Green Yellow | 57 | 64 | 7 | 5 | 6 |
| Yellow Green | 98 | 63 | 14 | 12 | 11 |
| Blue | 198 | 72 | 309 | 91 | 251 |
| Dodger Blue | 32 | 45 | 56 | 8 | 6 |
| Deep sky Blue | 19 | 68 | 3 | 6 | 7 |
| Purple | 89 | 83 | 136 | 21 | 88 |
| Red Purple | 36 | 51 | 24 | 17 | 16 |
| Magenta | 17 | 32 | 57 | 37 | 15 |
| Black | 13 | 64 | 12 | 213 | 47 |
| Dark Gray | 12 | 74 | 7 | 172 | 6 |
| Light Gray | 5 | 64 | 24 | 58 | 8 |
| Green Yellow | 16 | 36 | 7 | 62 | 3 |
| Total | 952 | 948 | 952 | 953 | 957 |
| STDEV | 60.5 | 16.7 | 88.3 | 53.5 | 93.2 |

V. CONCLUSION

This paper suggests color-music model to improve the effectiveness of general music retrieval and to assure perpetuity by new music. Further, suggesting new method to form feature vector from sound source and to apply it to Convolution Neural Network, it showed fruitful result. It is necessary to improve performance by increasing the number of sound source to learn color-music model and by changing the mode to select representative section of sound source in order to apply to the actual system later.

REFERENCE

- [1] Alfredo Raglio, Lapo Attardo, Giulia Gontero, Silvia Rollino, Elisabetta Groppo, and Enrico Granieri, Effects of music and music therapy on mood in neurological patients. World journal of psychiatry 5.1, pp.68-78 (2015).
- [2] Lee Jung Ho, Classical Music Marketing in the Age of Experience and Communication, Samsung Economic Research Institute Business notes No.142 (2012).
- [3] S. E. Palmer, K. B. Schloss, Z. Xu, L. R. Prado-Leon, Music-color associations are mediated by emotion, Proceedings of the National Academy of Sciences, Vol.110, No.22, pp.8836-8841 (2013).
- [4] J. M. Barbieri, A. Vidal, D. A. Zellner, The Color of Music: Correspondence through Emotion, Emperical Studies of the Arts, Vol.25, No.2, pp.193-208 (2007).
- [5] R. Bresin, What is the color of that music performance?, Proceedings of the International Computer Music Conference, pp.367-370 (2005).
- [6] C. B. Moon, H.S. Kim, H. A. Lee, B. M. Kim, Analysis of Relationships Between Mood and Color for Different Musical Preferences, Color Research & Application, Vol.39, No.4, pp.413-423 (2014).
- [7] M. Alex Syaekhoni, Development of Deep Learning Models for Multi-class Sentiment Analysis, Journal of Information Technology Services 16.4, pp.149-160 (2017).
- [8] Xin Liu, Qingcai Chen, Xiangping Wu, Yan Liu, Yang Liu, CNN based music emotion classification, arXiv: 1704.05665 (2017).