http://doi.org/10.17548/ksaf.2020.06.30.387

영상과 음악의 구성요소 간 상관관계 기반, 지능형 OST 매칭 자동화 프로그램 개발 연구

황 진 수 *
하 동 환 **
* 중앙대학교 창의ICT공과대학 연구원
** 중앙대학교 예술공학대학 교수

A Study on the Development of Intelligent OST Matching Automation Program based on the Correlation between the Components of Image and Music

Hwang, Jin Soo *
Har, Dong-Hwan **
* Researcher, College of ICT ,Chung-Ang University
** Professor, College of Art&Technology, Chung-Ang University

본 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A5A2A01031624)

** Corresponding Author : Har, Dong Hwan dhhar@cau.ac.kr

THE KOREAN SOCIETY OF SCIENCE & ART
한국과학예술융합학회

THE KOREAN SOCIETY OF SCIENCE & ART Vol.38(3)_Regular article or full paper
* Contribution : 2020.03.24_Examination : 2020.05.26_Revision : 2020.06.11_Publication decision : 2020.06.30

Contents

Abstract

This study was conducted to design a system called 'Intelligent OST matching'. The purpose of this study is to find out the relationship between image and music. For the configuration parameter of image, 'emotion', 'color' and 'pixel motion speed' were selected, and for music, 'mode','tempo', 'pitch', and 'loudness' were selected. Then I made program in order to analyze image and music and also made model to organize the results. As a result, configuration parameters of image and music has a relationship. For instance, pixel movement speed and brightness are related with tempo, and saturation is related with mode and pitch. Also, each parameters more complexly connected rather than a 1:1 response.

Key Words

Intelligent OST matching software, Image recognition, music components, image analysis

I. Introduction

Recently, with the development of media platform-based SNS service and the popularization of media production technology, the cases of producing and sharing images have been explosively increasing. According to a press release distributed by the Ministry of Education in 2019, In elementary school students[1], the 3rd place in the desired job ranking is 'creator'. Creator is a term that includes youtubers and streamers, and the above result indicates that many students are interested in making and sharing video. In addition, according to a report on the Me-media published by SK Securities in 2019[2], it is said that sales of MCN (Multi-Channel Network) companies that manage the contents creation and business of creators are rapidly increasing. The above statistics show that the demand for media platform-based SNS is increasing and has high potential in the future. In such a situation where video production and sharing are popularized, it is very important to insert background music suitable for the video to be produced. This is because appropriate background music of video can help the message or content to be conveyed more clearly or emphasized. However, there are many cases in which the use of inappropriate background music in many images makes the message rather obscure. In order to solve this problem, this study intends to find out the correlation between the configuration parameters of the image and the music that is the basis of the 'Intelligent OST matching program'; program that matches the music suitable for the image automatically. For example, if an amateur video producer wants to find music suitable for the video, he or she can be recommend suitable music and use it for the video by using this program. For this program, it is necessary to find an objective correlation between image and music. In order to find out the correlation between the image and the music, at first, this study chooses the configuration parameters of the image and the music. Each configuration parameters should be able to be summarized in quantitative manner and have the relationship.

The first configuration parameters of analysis is the emotion. Because, research and application of the connection with emotions are being conducted in both the image and music. In the case of music, many musicologists, led by Kate Hervner, studied

the associations. And, in the case of image, there are many APIs that can be easily used by general users in video. The second configuration parameters of analysis is the color. Because there are many analysis methods related to music, so the data can be obtained easily and quickly. The third configuration parameters of analysis is the pixel motion speed which shows the speed or dynamic of the image. It is related to the narrative of the image and closely related to the tempo of music. Therefore, this research tried to find the relationship by analyzing the relationship between music and pixel motion speed.

In conducting the research, the experiment was carried out in consideration of the weight between configuration parameters of music and image, rather than simply 1:1 matching. The images used in the analysis were selected for winning two or more related awards. For analysis, this research made a program using existing research and a new analysis tool to organize the analysis results.

II. Experiment

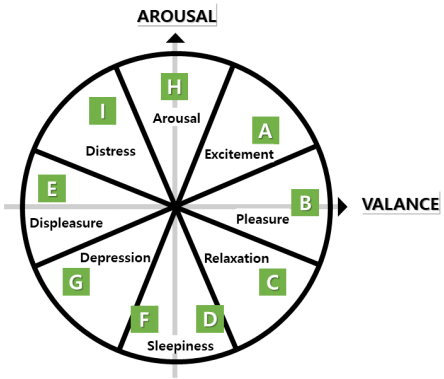
2.1 Preceding Studies on Music and Emotions

The study of music and emotions began with the study of Kate Hervner in 1935. The study was carried out by classifying 67 emotion-adjectives into eight and selecting subjects for music and appropriate adjectives. The result was very consistent regardless of musical knowledge or educational level. This can be summarized with following table

[Table-01] Hervner's study results table.[3)]

Emotions	solemn	sad	sentimental	serene
Mode	Major4	minor20	minor12	Major3
Tempo	Slow14	slow 12	slow16	sloe20
Pitch	low10	low 19	high6	high8
Emotions	graceful	happy	exciting	majestic
Mode	Major21	Major24	-	-
Tempo	fast6	fast 20	fast21	fast6
Pitch	high16	high6	low9	low13

In 1989, Russell made a new model using two-dimensional schematics of emotional adjectives. The horizontal axis (X-axis) is divided into Negative-Positive and the vertical axis (Y-axis) is expressed as Sleepiness-Energy according to the activity level. In Russell's model, 28 emotional adjectives were placed in eight groups. Subsequently, music educator Schubert presented a integrated model that combines the Hervner and Russell's emotional adjectives. The model categorized 46 adjectives into nine groups, which were found to be highly relevant to the previous model. This tool is also made in two dimensions, with each axis equal to Russell's model. The result can be summarized as following charts. An eight-part circle is group of sentiment adjectives categorized by Russell's method and the alphabet groups follow Schubert's method.



<Figure-01> Russell and Schubert's Classification[4)]

[Table-02] Schubert's Emotion Group[5)]

Group	Emotions
A(Bright)	Cheerful, Happy, Joyous
B(Lyrical)	Humorous, Light, lyrical
C(Calm)	Graceful, Relaxed
D(Dreamy)	Sentimental
E(Tragic)	Yearning
F(Dark)	Depressing, Melancholy
G(Majestic)	Heavy, Majestic, Scared
H(Dramatic)	Exciting, Passionate
I(tense)	Agitated, Restless

This study integrates Hervner's data into a two-dimensional model used by Russell and Schubert and uses the correlation between each axis and music. This method is similar to the method used in Meyers's previous studies(2007), and it was judged to be the most effective way to quantify and organize the configuration parameters of music.

2.2 Preceding Studies on Music and Color

Attempts to connect music with color began in ancient Persia. Among the ancient Persian philosophers, there was a record of one-to-one matching between sound and color, and similar attempts have been made several times from then on. Among them, Newton has spectroscopically analyzed solar light through prism and analyzed the sound from C to B according to the wavelength and color of light, from low frequency to high frequency color. In the case of the composer Skryabin, some attempts have been made to integrate music and art by assigning colors and meaning to the sound. For example, he inserted a part called 'light' into the score so that the color of the stage light changes according to the music. Like this, attempts to connect music and color can be said that they have continued from ancient times to modern times. In other words, from the past to the present, there is a perception that music and color are related in people.

A direct relationship of music and color can be found in the work of Karwoski and Henry S.Obert in 1938. According to the study, pitch is related to color. As the pitch goes up from low to high, the color reaction occurs from black to red to blue to gray to silver gray. In addition, the musician Christopher Ward presented color associated with the tone, and the French musician, Albert Lavignac, studied the color expression of musical instrument by mode(Major, minor).

The model used in this study is the color model presented in 'Basic Study for the Production of Color Music Image Scale' (Jung, Yoo Jung, 2014). This model is based on the Schubert's model mentioned in the previous research on music and emotions.

2.3 Preceding Studies on on music and pixel motion speed

The motion speed of each pixel, which is the minimum constituent unit of the image, is a criterion to determine how dynamical the image is. In other words, the faster a scene transits or a image moves, the more dynamic the image becomes. This can be found in works with music and movements such as ballet and Nonghak. In the case of Nonghak, performers turn their head faster matched to the tempo. Also in the case of Stravinsky's Spring Festival Ballet, dancers move faster and slower matched to the tempo. According to Kim's previous studies(2007) that analyzed Disney's Fantasia 2000, which visualized classical music, also found that the speed and dynamics of the song were visualized by changes in the speed and movement of the video.[6] In other words, the dynamics of video and the tempo of music are related and they can be said to have a positive correlation each other.

III. Research Method

3.1 Select image to analyze

In order to find the correlation between music and images, this study selected images to be analyzed. In the case of a film, selected film won a two or more music-related prizes at the film awards. In the case of a music video, it was selected based on the songs that won two or more image-related prize at the music awards. In the case of movies, the scene where the main theme of OST appeared was analyzed. And in the case of music videos, the entire contents of videos were analyzed. The videos and award lists are as follows.

①The Lord of the Rings: The Return of the King
·2014 Best Score Soundtrack For Visual Media (GRAMMY AWARDS)
·2014 Golden Globe Award for Best Original Score
·2013 Academy Award for Best Original Score
②Black Panther
·2018 Best Score Soundtrack For Visual Media (GRAMMY AWARDS)
·2018 Academy Award for Best Original Score
③The Grand Budapest Hotel
·2014 Best Score Soundtrack For Visual Media (GRAMMY AWARDS)
·2014 Academy Award for Best Original Score
④Bad romance - Lady Gaga

·2010 Best Music Video, Short Form (GRAMMY AWARDS)
·2010 Video of the year(MTV Video Music Award)
⑤Formation – Beyonce
·2016 Best Music Video, Short Form (GRAMMY AWARDS)
·2016 Video of the year(MTV Video Music Award)

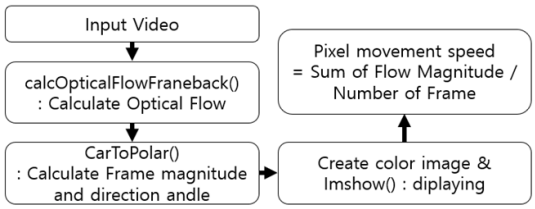
3.2 Create program to analyze image and music

(1) Image analysis method

For the image analysis, this study creates a new program based on Maeng's previous research(2019) In the case of emotion, the analysis was conducted by using Microsoft's 'Azure Video Indexer.' The application analyzes the image and provides information such as keyword extraction, face recognition, and emotion analysis results. This study utilizes the emotion information derived from the results analyzed by this applications at the part of emotion analyzing.

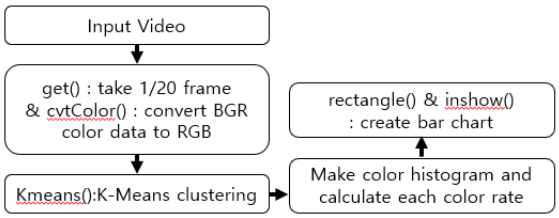
The pixel motion speed was analyzed by using 'Dense Optical Flow'. Dense Optical Flow is a motion measurement algorithm between two consecutive frames based on Polynomial Expansion. Optical flow refers to the movement pattern of an object, surface, or edge in an image and is caused by the movement of an object or camera, which can be interpreted as the movement velocity distribution of the image brightness pattern.

In the program produced in this study, based on the result value of 1.55, if it is higher than 1.55, that analyzed image is determined as a high dynamic image. If not, it is determined as a lower dynamic image. The criteria 1.55 was determined based on the mean value of pixel motion speed extracted from the this study and previous study. In the previous study, the same analysis method used in this study was used to calculate the pixel motion speed for multiple randomly selected images, mean value of pixel motion speed is 1.50175 in the random samples. It can be said that the images from the previous study were selected at random, so it is a more common result. The detailed process of the program is as follows.



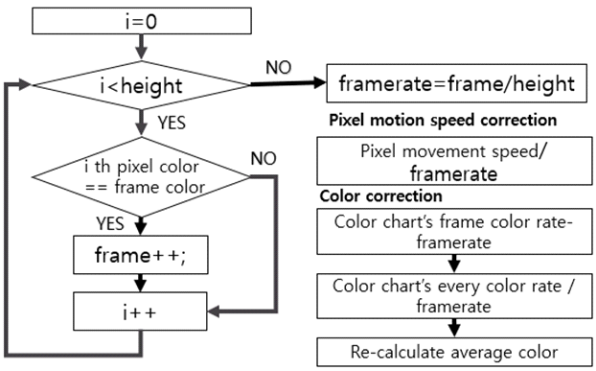
<Figure-02> 'pixel motion speed' calculation algorithm[7)]

Color was analyzed using Image Histogram. First, 1/20 of the entire frame of the image was taken. Then, K-means clustering was performed to make n clusters for learning. Based on the learning results, the average colors and their proportions were obtained. In this study, 10 clusters were created. The detailed process of the program is as follows.



<Figure-03> 'color' calculation algorithm[8)]

Among the analyzed images, there are images with black frames. In this case, the result was corrected by the ratio of the frame part in the overall result. The resultant was corrected by reading the pixels in vertical lines to obtain the frame rate and applying them to the pixel motion speed and color. The correction algorithms is as follows.



<Figure-04> Correction algorithm

(2) Music analysis method

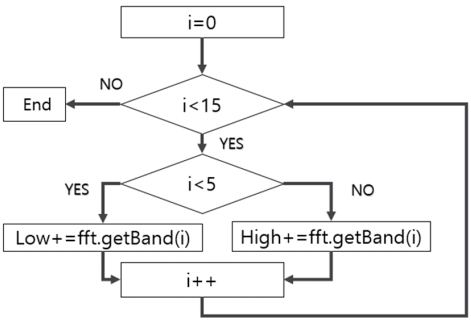
Music analysis was based on musical scores. Score which officially provided by film or music producers, or music scores containing original sound information were selected.

Mode was determined major and minor according to the main melody's mode.

Loudness is classified by dynamic marking. If there are many melody above mp (mezzo piano), the music was determined as a loud music. If not, it was determined as a quiet music.

Tempo is classified by BPM. If it is higher than 120 BPM it was determined as a fast music, if not, it was determined as slow music.

Pitch is classified by 'middle C' note. If most of the notes are higher than 'middle C' note, music was determined as high-pitched tone, if not, it was determined as low-pitched tone. For vocal music, the vocal range was analyzed by using the fft.getBand (int index) function from Processing3's Minim library. This function shows the result of FFT analysis, and frequency is divided by index variables. For the transliteration analysis, this study created the 'Low' and 'High' variables. If the index variables of fft.getBand (int index) function is less than 5, it is recognized as a low note and the corresponding spectrum value is stored in the 'Low' variable. If not, corresponding spectrum value is stored in the 'High' variable. This process is repeated until the end of the song, and the pitch is determined by a larger variable.. The detailed algorithm of the program is as follows.



<Figure-05> Pitch calculation algorithm

3.3 Create tools to organize analysis result

In order to organize the analysis, this study create new analysis model based on Russell and Schubert's two-dimensional model. Based on emotion, color, and pixel motion speed, configuration parameters of image and music are summarized in a two-dimensional model.

In the case of emotions, this research finds the relationship between music configuration parameter and emotions by applying Hervner's result to Russell and Schubert's Two-dimensional model. The results are summarized as follows.

The scales of positive and negative corresponding to the X-axis of the 2D axis are related to mode, pitch, and loudness. If it is closer to the 'Positive', Major, loud and high-pitched tone music is suitable. If it is closer to the 'Negative', minor, quite and low-pitched tone music is suitable.

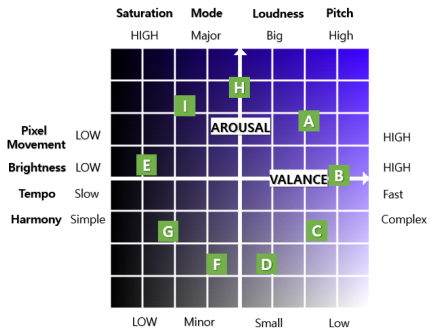
The scales of activeness and sleepiness, which correspond to the Y-axis of the 2D axis, were related to the complexity of tempo and chord. If it is closer to the 'Activeness', faster tempo and simple chords music is suitable. If it closer to the 'Sleepiness', slower tempo and complex chords music is suitable.

Next, the colors are summarized. The color of the color model, based on Schubert's two-dimensional model, was analyzed for each quadrant. The result is shown below.

Saturation and brightness are related to the configuration parameter of music.

In the case of saturation, it is matched with the dynamic and sleepiness corresponding to the Y-axis, and in the case of brightness, the positive and negative scales corresponding to the X-axis are matched.

Finally, the pixel motion speed is placed on the X axis in relation to the tempo. The model is summarized as follows.



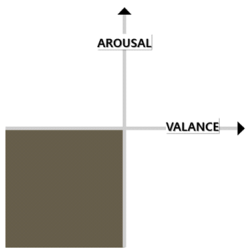
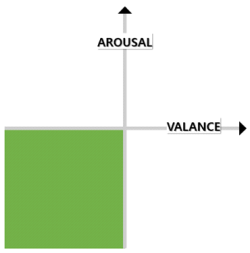
<Figure-06> Analysis model of image and music

This research organize each result. If the results are the same, it can be said that music is used appropriately for the image.

And this study consulted experts for objective verification of this research method. As a result, Consultation was given on whether the Configuration Parameters were well selected and whether the selected Configuration Parameters measurement method(program) was properly designed. As a result, configuration Parameters and their measurement methods have been selected and designed accordingly.

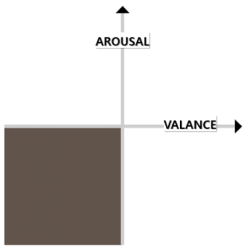
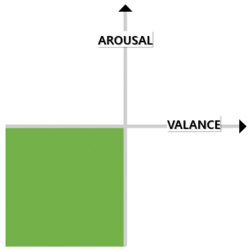
IV. Result

[Table-03] Analysis result of The Lord of the Rings

The Lord of the Rings: The Return of the King -<The Return of the king>			
Image analysis result		Image analysis result	
			
emotion	Sad	mode	D Major
saturation	42	loudness	Small
brightness	88	tempo	Slow
pixel motion	1.6	pitch	Low

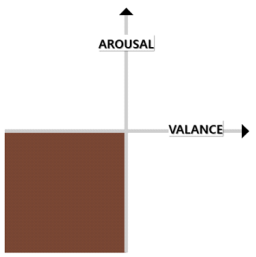
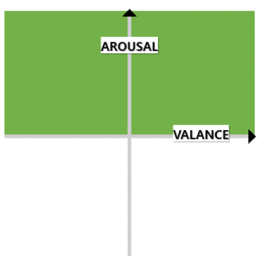
As a result of analyzing 'The Return of the king', the theme song of the ending part in 'The Lord of The Rings', main emotion is sadness, and the brightness and saturation of the average color are very low. The pixel motion speed is 1.6, which is slightly higher than the reference value. Music is Major mode music, but it is characterized by low pitches, slow tempo, and quiet. Based on the above results, it was concluded that the overall harmony between the theme song, 'The Return of The king' and the movie was well achieved.

[Table-04] Analysis result of Black Panther

Black Panther -<Wakanda>			
Image analysis result		Image analysis result	
			
emotion	Negative	mode	C minor
saturation	32	loudness	Loud
brightness	85	tempo	Slow
pixel motion	1.4	pitch	Low

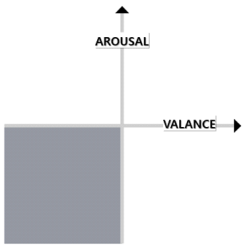
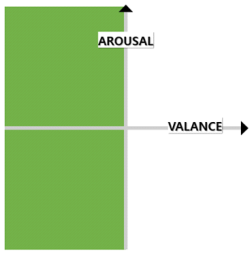
As a result of analyzing 'Wakanda', the theme song of the film 'Black Panther', main emotion is negative, and the brightness and saturation of the average color are very low. The degree of motion of the pixel is 1.4, which is lower than the reference value. Music is minor mode music, and it is characterized by low pitches, slow tempo, and loud. The theme song 'Wakanda' also showed a good harmony.

[Table-05] Analysis result of The Grand Budapest Hotel

The Grand Budapest Hotel -<Mr. Moustapha>			
Image analysis result		Image analysis result	
			
emotion	Sad	mode	C Major
saturation	98	loudness	Small
brightness	83	tempo	Fast
pixel motion	0.68	pitch	High

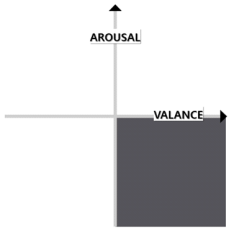
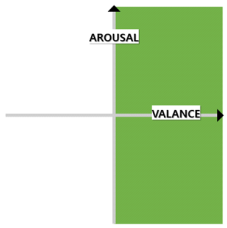
As a result of analysis of 'Moustapha', the theme song of the film 'Grand Budapest Hotel', main emotion in the image part is sadness, and the average color brightness and saturation are low. The degree of motion of the pixel is 0.68, which is lower than the reference value. However, in the case of music, the fast tempo with high pitches notes were played mainly in major music. The image model and the music model showed different results compared to each other. This can be interpreted as follows. Although the song is Major in music, Major and minor melodies cross because of accidental mark. This suggests that the music to be used for 'Intelligent OST matching program' is better suited to music with a consistent mode without many accidental mark and modulation.

[Table-06] Analysis result of Bad Romance music video

Lady Gaga - <Bad Romance>			
Image analysis result		Image analysis result	
			
emotion	-	mode	A minor
saturation	6	tempo	Slow
brightness	120	pitch	High
pixel motion	1.48		

As a result of analyzing Lady Gaga's <Bad Romance>, main emotion doesn't appear. So images were analyzed based on the color and pixel motion speed except emotion. The saturation is very low and the brightness is medium level. And the pixel motion degree is 1.48, which is lower than the reference value. Music is minor mode music, and it is characterized by high pitches, slow tempo, The music video can be said to be well made according to the characteristics of the music.

[Table-07] Analysis result of Formation music video

Beyonce - <Formation>			
Image analysis result		Music analysis result	
			
emotion	Positive	mode	F minor
saturation	9	tempo	Fast
brightness	83	pitch	High
pixel motion	1.98		

As a result of analyzing Beyonce's <Formation> music video, main emotion is positive in the video part, and the saturation and brightness are low. The pixel motion was very high at 1.98. Music is minor mode music, and it is characterized by high pitches, fast tempo, The music video can be said to be well made according to the characteristics of the music.

V. Conclusion

With the popularization of media platform-based SNS service and the development of media production technology, people can easily produce and share videos. This tells us that image has become a universal means of communication. Music in video is a tool that effectively conveys the message you want to deliver. However, selecting music suitable for video can be very difficult and cumbersome for creators. Therefore, this study was conducted as a preliminary study for producing a system called 'Intelligent OST Matching program' which analyzes the produced images and automatically recommends musics that are suitable to them. For this purpose, three configuration parameters (emotion, color, and image motion) were selected. To analyze image and music, this study created new programs. In addition, to organize the results, this study made analysis model. In the case of the images used for analysis, images which won two or more prizes in the related award were selected, and in the case of music, this study analyzed main theme in films and entire song in music videos. As a result of the verification, 4 out of 5 images showed agreement, which means

that a significant relationship between image and music was found. Subsequent research will require more viewer-centered verification through the survey method. It is also expected that more analysis will be conducted based on popular videos on TV programs or media platform-based SNS. In addition, as Korean video and music contents are gaining global popularity, we expect to develop an integrated OST matching program specialized in certain countries based on the methods used in this research. This study examined the association between image and music as a preliminary step for intelligent OST matching program. It is hoped that this study can be used appropriately to develop intelligent OST matching program that can automatically match appropriate music for individual video producers.

Reference

[1] Hevner. K,"Experimental studies of the elelments of expression in musinc", The American journal of psychology, 1936.

[2] Jung, Jinah "A Study on Relationship between Moving Image and Musical Pattern,Sungkyunkwan University, 2018.

[3] Jeong, Yoojeong., "A foundational study for developing colour-music image scale", Graduate School of Hongik University-Jeong, 2014.

[4] KIM, Eunju, "The Study on Correlation between Music and Image by Analyzing Music in <Fantasia 2000>",The Graduate School of Music in Modern Media, Ewha Womans University, 2008..

[5] Kim,Sangho. Kim,Sungtak. Kwon,Suk-bong. Ji,Mikyong. Kim,Hoirin. Yoon,JeongHyun. Lee,Hankyu."Automatic Music Recommendation System based on Music Characteristics", The HCI Society of Korea, 2007.

[6] Lee,Do Kyoung. Kim,jun,"Interdependence of Images and Music Combined by Sharing the Identical Properties - Based on the Movie 〈Interstellar, 2014〉 " The Korea Contents Society, 2019.

[7] Ministry of Education ,Career Education Policy Division ,"Announcement of 2019 Elementary and Secondary Career Education Survey",2019.

[8] Maeng, Chae Jung , Har, Dong-Hwan, "A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching", KOREA SCIENCE & ART FORUM, 2019.

[9] Owen Craigie Meyers, "A Mood-Based Music Classification and Exploration System", MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2007.

[10] Russell J.A, "Circumplex Model of Affact" Canada, Journal of personality and social psychology University of British Columbia, Vancouver, , 2010.

[11] Schubert.E ,Measuring emotion continuously : Val idityand Reliability of the two - dimen sional emotion - space,Australian journal of Psychology,1999.

[12] Schubert.E , Update of the Hevner adjective check list , Univerdity of new south wales,2003.

[13] Yoon,Hyukjin, Park,Chansol and Lee,Sojoong,"AIndustry Analysis :The fast-growing me-media industry, SK Securities.Co , 2019.

Endnote

1) Ministry of Education ,Career Education Policy Division ,"Announcement of 2019 Elementary and Secondary Career Education Survey", 2019.

2) Yoon,Hyukjin, Park,Chansol and Lee,Sojoong,"AIndustry Analysis :The fast-growing me-media industry ", SK Securities.Co , 2019.

3) Owen Craigie Meyers, "A Mood-Based Music Classification and Exploration System", MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2007.

4) Jeong, Yoojeong., "A foundational study for developing colour-music image scale", Graduate School of Hongik University-Jeong, 2014.

5) Jeong, Yoojeong., "A foundational study for developing colour-music image scale", Graduate School of Hongik University-Jeong, 2014.

6) KIM, Eunju, "The Study on Correlation between Music and Image by Analyzing Music in <Fantasia 2000>",The Graduate School of Music in Modern Media, Ewha Womans University, 2008.

7) Maeng, Chae Jung , Har, Dong-Hwan, "A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching", KOREA SCIENCE & ART FORUM, 2019.

8) Maeng, Chae Jung , Har, Dong-Hwan, "A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching", KOREA SCIENCE & ART FORUM, 2019.