



영상구성 파라미터 기반 지능형 OST 매칭 프로그램 개발 연구*

조 두 희**

중앙대학교 첨단영상대학원 박사후 연구원

하 동 환***

중앙대학교 예술공학대학 교수

I. 서론

1. 연구 목적

최근 급격한 관련 기술의 발전과 더불어 다양한 종류의 영상 콘텐츠가 생산, 소비되고 있다. 과거 고가의 하드웨어와 숙련된 기술을 가져야만 작업할 수 있었던 영상제작 시대에서 벗어나 지금은 특별한 기술이나 장비 없이도 누구나 쉽게 영상콘텐츠를 제작할 수 있는 사용자 창작 콘텐츠(User Created Contents)시대로 진입하였다. 이렇게 개인들이 전문기술과 장비 없이 수준급의 영상콘텐츠를 제작할 수 있게 되면서 참여형 콘텐츠 기반의 문화가 활성화 되고 있고, 산업적으로는 소규모 영상콘텐츠 분야의 창업 또한 활발히 이루어지면서, 다양한 형태의 콘텐츠공급자(Content Provider)들이 증가해 영상콘텐츠 산업의 질적, 양적 확대를 이끌고 있다.

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

** 제1저자, 중앙대학교 첨단영상 대학원 디지털과학사건연구실 박사후 연구원

*** 교신저자, 중앙대학교 예술공학대학 교수

그러나 폭발적으로 늘어난 영상제작 기회에 비해 영상 결과물의 완성도에 큰 영향을 미치는 적절한 배경음악(OST-Original Sound Track)의 선정에는 많은 어려움이 존재하고 있다. 고비용이 소요되는 음악 전문가와 협업이 어려운 개인 제작자들의 경우 영상과 어울리지 않은 OST를 삽입하여 전체적인 영상 수준의 저하는 물론 영상 콘텐츠를 통해 전달하려고 하는 메시지가 희석되는 경우가 빈번하게 발생하고 있는 것이 현실이다.

OST는 영상콘텐츠의 문맥과 비례하여 기본적인 정보와 추가적인 정보를 제공하는 기능을 가지고 있다, 즉 분위기를 조성하고 미학적 에너지를 추가하는 기본적인 기능, 그리고 커뮤니케이션을 강화하고 영상 방향력의 영역에 리듬구조를 보충하는 추가적인 기능을 수행할 수 있다.

본 연구에서는 제작된 영상이 가지고 있는 Scene의 구성요소들(Scene Design Factor - 시간, 동작, 빈도, 색상, 음향, 배경, 조명, 인물의 감정상태 등) 각각의 특징을 자동으로 분석하여 각 파라미터들과 조화를 이루는 음악의 구성요소들(선율, 리듬, 화성, 형식, 빠르기, 음색 등)을 가진 최적의 OST를 자동으로 선정, 영상 제작자에게 추천할 수 있는 지능적 배경음악 추천 프로그램을 개발하고자 하였다.

2. 연구배경 및 방법

제작된 각각의 영상의 구성요소들과 어울리는 배경음악을 찾기 위해서는 영상과 음악 각각의 특성들을 찾아내고 수치화하여 상호 비교할 수 있어야 한다.

맹채정, 하동환(2019)은 영상과 배경음악의 적합도를 측정할 수 있는 요소를 선정하고 수치화하여 영상과 음악의 정량적 비교방법을 구현하였다¹⁾. 이 과정에서 영상의 특성을 ‘감정’, ‘픽셀 움직임’, ‘색상’, ‘등장인물’로 선정하고 적절한 함수를 활용하여 프로그램을 제작, 각각의 데이터를 추출하였다. 감정을 위하여 사용된 함수는 ‘Azure Video Indexer’ 함수를 사용하였고, 픽셀 움직임은 ‘Dense Optical Flow’, 색상 분석을 위해서는 ‘Image Histogram’, 등장인물을 위한 데이터 추출은 CNN(Convolutional Neural Network)을 사용하였다. 이 함수들을 탑재한 프로그램은 영상을 자동적으로 분석하여 음악과 비교할 수 있는 단서를 만든다.

황진수, 하동환(2020)은 지능형 OST 매칭을 위하여 영상의 구성요소를 감정, 색

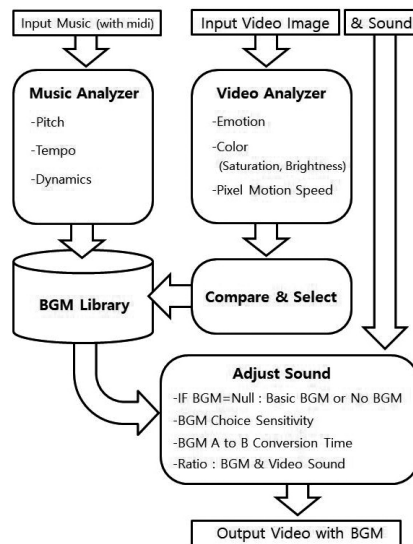
1) Maeng, Chae Jung, Har, Dong-Hwan, "A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching", KOREA SCIENCE & ART FORUM, v37, no.3, 2019, pp.125-134.

채, 움직임으로 특징하고 음악의 구성요소는 조성, 템포, 셈여림, 음고, 음대역 등으로 특징하여 각 요소간의 매칭을 시도하였다²⁾. 이것은 Russell의 2차원 모델에 대입하여 수치적 비교를 가능하게 하였으며 판단근거를 제시하였다³⁾.

본 연구에서는 위의 선행 연구들을 기반으로 영상에서 다섯 가지 구성요소를 추출하고 이를 기반으로 배경 BGM 라이브러리에서 이와 가장 상응하는 구성요소들을 가진 음악을 선별하여 영상에 실시간으로 적용해 보는 지능형 매칭 프로그램을 제작하였다.

II. 본론

1. 프로그램의 구조



<그림 1> Program structure

프로그램의 기본 구조는 [그림 1]과 같이 크게 5개의 블록으로 이루어진다. 프로그램 동작은 사전 BGM 라이브러리를 만드는 단계로 시작한다. 입력된 음악 소스에 대하여 ‘입력음악 분석(Music Analyzer)’ 과정을 통해 구성요소별, 특징별로 분리하

2) Hwang, Jin Soo , Har, Dong-Hwan, "A Study on the Development of Intelligent OST Matching Automation Program based on the Correlation between the Components of Image and Musi", KOREA SCIENC & ART,38(3), 2020, pp.387-396.

3) Russell J.A, "Circumplex Model of Affect" Canada, Journal of personality and social psychology, 96(6), 1980, pp.1161-1178.

여 'BGM 라이브러리(BGM Library)'를 구축한다. 이후 대상 영상에 대하여 '입력 영상 분석(Video Analyzer)'을 수행하고 찾아낸 영상의 특징을 '분석결과를 이용한 배경음악 찾기(Compare & Select BGM)'를 통하여 BGM 라이브러리에서 가장 상관관계가 높은 구성요소들을 가진 음악을 찾아 보여준다. 최종적으로 '기타 소리 설정(Adjust Sound)'에서 소리비율, 변환시간 등을 조절하고 영상과 함께 출력한다.

2. 입력 음악 분석 기제 및 BGM 라이브러리 제작

음악은 황진수, 하동환(2020)⁴⁾의 연구에 따라 '음고', '템포', '셈여림'으로 나누어 분석하였으며 분석방법은 Zheng Jinag(2019)⁵⁾가 제시한 MIDI파일의 멜로디, 베이스, 구조 분석방법을 활용하였다. 분석결과는 영상의 요소들과 비교하기 위하여 각 수치들을 10단계 변환하였다. 입력 MIDI 파일은 실제 악보 매 단락에 해당하는 위치에 마커를 등록, 동기화 하고 단락별 특성 분석을 원활히 진행되도록 하였다.

음고는 주 멜로디를 Processing Minim 라이브러리의 `fft.getBand(int index)`함수로 분석하여 얻으며, 푸리에변환 결과로 얻는 index를 이용하여 '가온다'음 C4를 기준으로 '고음'으로 5단계, '저음'으로 5단계를 구분하여 총 10단계로 나누었다. 매 단락을 반복하여 측정하며 평균 결과값이 큰 쪽으로 판단하였다. 템포는 BPM을 사용하여 40부터 184까지 구간을 10단계로 나누고 빠르기를 수치화 하였다. 셈여림은 연주자의 주관적 해석에 따라 다를 수 있지만 멜로디에 대한 전체음악의 강도 변화를 고려하여 셈여림 단계를 규정하였고, 가장 여리게에 해당하는 '피아니시시모'부터 가장 세게에 해당하는 '포르티시시모'까지 2부터 9까지 설정하여 총 8단계로 구분하였다.

단락별로 분석된 음악파일은 분석결과가 저장된 태그정보와 함께 별도로 BGM 라이브러리 명세서에 기록되도록 하였으며 MIDI파일과 연동시 참고자료로 사용되도록 하였다. 분석 및 규정이 끝난 특정구간 MIDI파일은 저장하는 과정에서 파일명에 단락정보를 추가시키므로 호출과정에서 악보의 위치를 확인할 수 있게 하였다.

4) Hwang, Jin Soo, Har, Dong-Hwan, "A Study on the Development of Intelligent OST Matching Automation Program based on the Correlation between the Components of Image and Musi", KOREA SCIENC & ART, 38(3), 2020, pp.387-396.

5) Zheng Jiang, "Automatic Analysis of Music in Standard MIDI Files" Music and Technology Carnegie Mellon Univ., May 2019.

3. 입력 영상 분석 기제

입력 영상의 분석을 위하여 맹채정, 하동환(2019)의 연구를 사용하여 ‘감정’, ‘움직임’, ‘색채’ 등을 분석에 활용하였다⁶⁾.

감정을 판단하기 위하여 Microsoft가 개발한 ‘Azure Video Index’를 이용하여 얼굴 인식, 감정 분석 등을 결과로 전달받는다. 많은 감정 요소를 ‘부정(Negative)적 감정’과 ‘긍정(Positive)적 감정’으로 구분하기 위하여 ‘Happiness’, ‘Surprise’값의 평균과 ‘Anger’, ‘Contempt’, ‘Disgust’값의 평균비율을 계산하여 10단계로 나누었다. 또한, ‘동적(Activeness)인 감정’과 ‘정적(Sleepiness)인 감정’을 구분하기 위하여 ‘Happiness’, ‘Surprise’, ‘Anger’값과 ‘Sadness’, ‘Neutral’값의 평균의 비율을 계산하여 10단계로 나누었다.

색채는 Histogram 함수를 이용하여 채도(Saturation)와 밝기(Brightness)로 구분하여 분석하였다. 분석을 위하여 입력 영상을 1.5초 단위로 프레임을 누적하고 평균화된 이미지로 만들도록 하였다. 이것을 기초하여 K-평균 군집화(K-means clustering)를 하고 n개의 clusters를 만들어 학습하도록 하였다. 이 결과는 평균 색채와 그 비율을 측정할 수 있도록 하였고 생성된 clusters가 50개가 모이면 대표 clusters를 새롭게 생성시켜 분석에 활용하도록 하였다.

채도는 생성된 clusters의 RGB컬러 비율을 이용하며 낮을수록 정적(Sleepiness)인 영상으로, 높을수록 동적(Activeness)인 영상으로 판단하도록 하였다. 비율수치는 0.0부터 1.0까지 구간만 사용하며 10단계로 나누어 채도를 표시하도록 하였다.

밝기는 생성된 clusters의 RGB컬러를 Lab컬러로 바꾸고 L(Luminance)값의 높고, 낮음에 따라 부정(Negative)과 긍정(Positive)의 척도로 구분하였다. 0%부터 100%사이에서 생성된 수치는 10단계로 나누어 밝기를 표시하도록 하였다.

움직임은 카메라 워킹, 피사체의 움직임의 차이를 구분하지 않으며 ‘Dense Optical Flow’를 이용하여 분석하였다. 이 분석 도구에 기반한 Polynomial Expansion 함수는 이어지는 프레임들의 유사성을 판단하고 움직임의 범위와 정도를 측정할 수 있다. 움직임의 근거는 영상의 등장인물이나 경계선의 움직임 패턴을 의미하며 이미지 밝기 패턴의 움직임 속도 분포로 해석할 수 있다. 제작하는 프로그램의 초기 시작값은 1.55로 하며 0.0부터 3.0까지 범위를 10단계로 분류하여 움직임

6) Maeng, Chae Jung , Har, Dong-Hwan, “A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching”, KOREA SCIENC & ART FORUM, v37, no.3, 2019, pp.125-134.

이 많고, 적은 영상을 판단하도록 하였다.

4. 분석결과 시각화 및 배경음악 선택

분석된 입력 영상 요소들의 결과를 토대로 BGM 라이브러리 음악파일들의 분석 요소와 비교하여 가장 유사한 값의 음악파일을 찾게 하였다. 영상과 음악의 분석 요소들은 다음 [표 1]과 같이 짝을 이루어 비교한다. 영상의 감정, 색채, 움직임 등의 요소를 음악과 비교하기 위하여 Meyers(2007)⁷⁾의 비교방법과 김은주(2008)⁸⁾의 연구 등을 참고하였다. 또한 각 요소들의 분석과정에서 1부터 10단계로 분류하여 상호 비교가 용의하게 하였다.

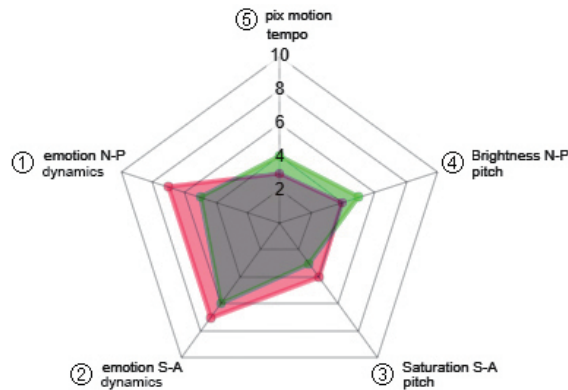
	영상 분석 요소	음악 분석 요소
①	감정 : 부정(Negative)-긍정(Positive)	샘여림(dynamics)
②	감정 : 정적(Sleepiness)-동적(Activeness)	샘여림(dynamics)
③	색채(채도) : 정적(Sleepiness)-동적(Activeness)	음고(pitch)
④	색채(밝기) : 부정(Negative)-긍정(Positive)	음고(pitch)
⑤	움직임 : 느리게(Slow)-빠르게(Fest)	템포(tempo)

<표 1> Comparing elements of analysis between image and music.

영상과 선택된 배경음악의 분석결과와 [그림 2]와 같이 방사형 5각 그래프에 표시하여 시각적으로 차이를 한눈에 확인할 수 있게 하였다. 5각 그래프의 각 꼭지점은 [표 1]의 비교요소를 동시에 표시할 수 있으며 영상은 초록색 5각형으로 배경음악은 붉은색 5각형으로 표시하여 쉽게 구분할 수 있도록 하였다. 유사한 값의 배경음악을 찾기 때문에 두 도형은 매우 유사한 형태를 띠게 되며 BGM 라이브러리에 배경음악이 많을수록 두 도형의 모양은 유사해 질 수 있는 확률이 높아진다.

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

8) 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.



<그림 2> A radial pentagonal graph for comparison.

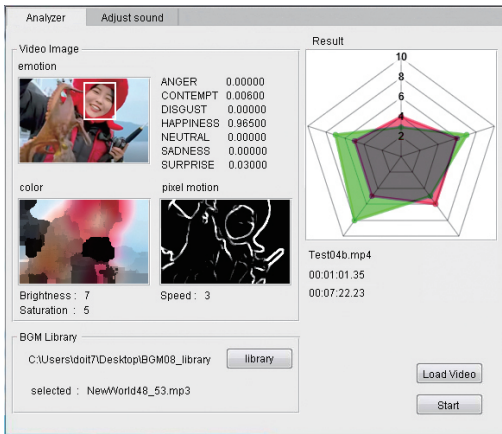
5. GUI 및 기타설정

앞서 제작한 프로그램은 [그림 3]과 같은 GUI를 통하여 동작하도록 하였다. 영상 분석 결과, 선택된 배경음악, 비교결과 5각 그래프 등을 표시한다. 또한 배경음악을 출력할 때 다음의 네 가지 설정을 사용자가 지정할 수 있도록 만들었다.

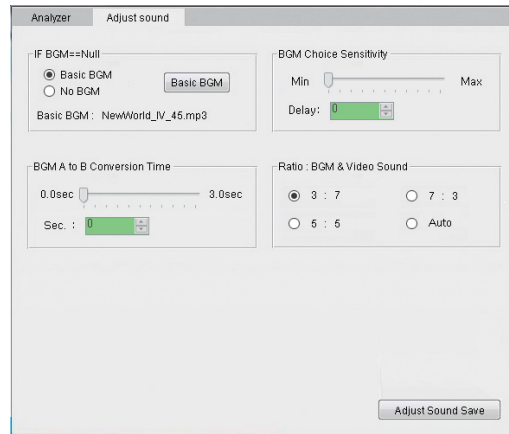
첫 번째는 BGM 라이브러리로부터 적당한 파일을 찾지 못한 경우이다. 이 경우 배경음악 없이 비디오-사운드만 사용하거나, 기본 배경음악을 무한반복으로 사용하는 방법 중 한 가지를 선택하여 사용할 수 있도록 하였다.

두 번째, ‘배경음악 선택 민감도 조절(BGM Choice Sensitivity)’이다. 입력영상은 주제와 상관없는 여러 가지 요소가 비주기적으로 화면에 나타나고 내용과 상관없는 움직임, 표정 등으로 인하여 배경음악의 분위기가 산발적으로 변할 수 있다. 이러한 문제를 감소시키기 위하여 배경음악 선택 민감도 조절 시간을 설정할 수 있게 하였다. 이것은 배경음악의 최소 재생시간을 정하고 그 사이 반복되는 영상분석 누적 값을 평균하여 배경음악을 결정한다.

세 번째, 현재 재생하고 있는 배경음악에서 새롭게 선택된 배경음악으로 바뀌는 순간 갑작스런 변화가 일어나지 않도록 하기 위하여 소리의 페이드-인(Fade-in), 페이드-아웃(Fade-out) 시간 조절기능을 추가하였다. 이 기능을 통하여 배경음악이 바뀔 때 음악이 자연스럽게 이어지도록 하였다.



<그림 3> Image, music analysis, and comparison results.



<그림 4> Adjustment window for sound ratio and other settings.

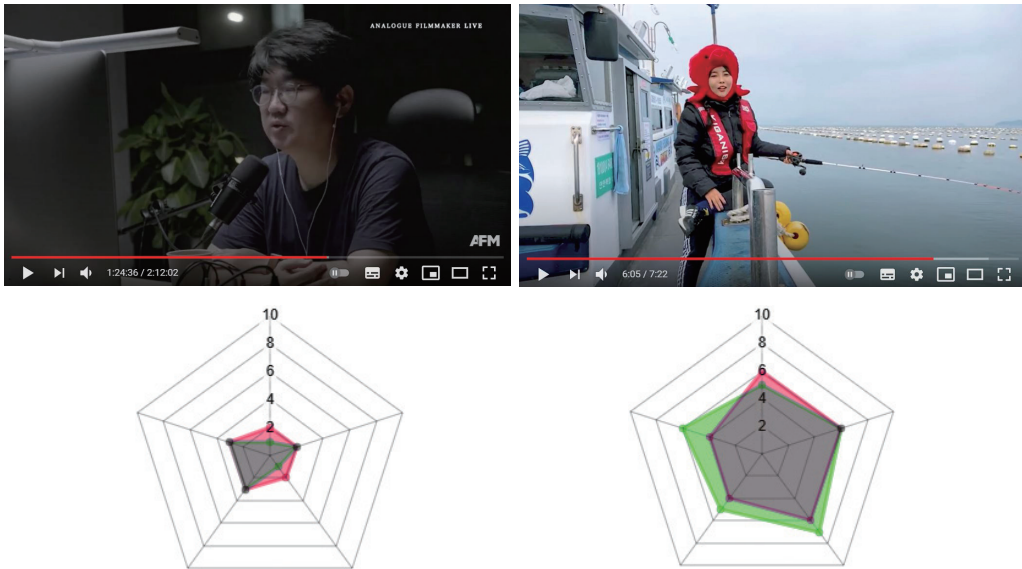
마지막으로 선택된 배경음악과 함께 비디오-소리(Video-sound)를 적당한 비율로 혼합할 수 있는 기능을 추가하였다. 스튜디오에서 대화로 방송을 이끌어가는 경우 비디오-소리의 크기 비율이 높아야 하지만 활동적인 스포츠나 야외활동을 방송하는 경우 상대적으로 배경음악의 크기 비율이 더 높아야 한다. 이러한 모든 경우는 사전에 소리비율을 3:7, 5:5, 7:3 등으로 설정하여 최적의 음향비율을 구현할 수 있다. 또한 자동(Auto)기능을 이용하여 비디오-소리의 유무에 따라 배경음악과의 비디오-소리의 비율이 자동으로 조절할 수 있게 하였다. [그림 4]는 소리비율 및 기타 조절 창을 보여준다.

6. 프로그램 적용

제작한 프로그램의 동작을 확인하기 위하여 다양한 영상 파일을 대상으로 테스트하였다. 실험에 사용한 영상은 빠른 화면전환 등이 없는 비교적 정적인 형태와 활발한 야외활동을 포함하는 동적인 영상 등을 다양하게 사용하였으며 간단한 캐릭터 애니메이션도 테스트에 적용해 보았다. BGM 라이브러리는 드보르작 교향곡 9번 4악장 '신세계로부터'(Dvorak, Symphony No.9, 4th 'from the New World)를 활용하여 구축하였다. 이 곡은 E단조 4/4박자, 소나타 형식으로 힘차고, 빠른 느낌, 고요하고, 차분한 선율 등을 함께 포함하고 있으므로 여러 단락으로 나누어 다양한 영상의 배경음악에 적용할 수 있도록 하였다.

아래는 적용 결과들의 예이다. 정적인 영상에 적용해 본 예인 [그림 5]⁹⁾에 대한 프로그램의 분석결과를 방사형 5각 그래프로 보여주고 있다. 상응하는 배경음악은 NewWorld65_80.mp3 와 NewWorld102_107.mp3 등을 번갈아가며 선택하여 영상에 대한 느리고 차분한 분위기를 잘 표현하였다.

두 번째 영상은 20대 여성이 시청자와 대화하며 낚시하는 모습의 영상을 사용하였다. 촬영시간이 낮 시간, 야외이므로 밝고, 높은 채도의 영상이다. 인물의 표정은 자주 바뀌지만 Happiness 점수가 대체적으로 높게 받았다. 하지만 측면이 촬영되는 빈도가 높아 얼굴을 인식하지 못하는 경우가 많았다. 이러한 경우 이전 Emotion 값을 그대로 사용하므로 배경음악 선택에 별다른 문제를 만들지 않는다. [그림 6]¹⁰⁾의 방사형 5각 그래프로 분석결과를 확인할 수 있다. 상응하는 배경음악은 NewWorld106_109.mp3 와 NewWorld128_144.mp3, NewWorld172_183.mp3 등을 선택하여 활동적이며, 경쾌한 분위기를 표현하였다.



<그림 5> 'analogue filmmaker' analysis result <그림 6> '흔치않은여자 예라니TV' analysis result

세 번째는 운동을 하는 활동적인 영상을 사용하였다. 특별히 대화하는 장면은 없으며 얼굴 인식률이 낮아 영상 초반 인식된 얼굴의 측정값을 그대로 유지하였다. 콘텐츠 특성상 고속의 움직임으로 인하여 매우 높은 Pixel motion값을 얻는다. [그림-7]¹¹⁾은 적용한 영상의 모습을 보여주며, 5각 그래프는 그 분석결과이다. 상응하

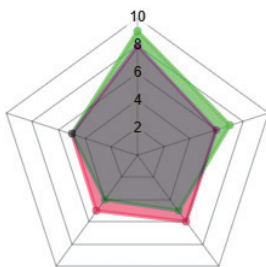
9) <www.youtube.com/watch?v=VVTUXrz0iJU>, <검색일 2021년11월30일>

10) <www.youtube.com/watch?v=lxnxxHDji2E&t=325s>, <검색일 2021년11월30일>

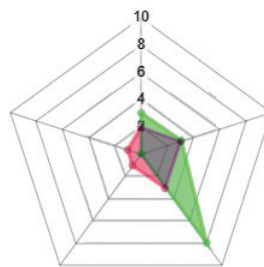
11) <www.youtube.com/watch?v=vSBcrmx4aFw>, <검색일 2021년11월30일>

는 배경음악은 NewWorld44_49.mp3와 NewWorld94_99.mp3 등을 번갈아가며 선택하여 빠르고 경쾌한 분위기를 잘 표현하였다.

마지막으로 애니메이션 <판타지아2000>의 ‘마법사의 제자’부분에 적용해 보았다. 이 영상의 채도가 높으며 강한 콘트라스트의 영상이다. 또한 움직임 속도가 빈번하게 변하고 의인화된 캐릭터 얼굴표정을 Azure 함수가 인식하지 못하는 특징이 있었다. 이러한 경우 대부분 배경음악을 찾지 못했을 때 선택되는 Basic BGM을 선택하였다. 결과적으로 Pixel motion값을 제외한 다른 요소에 대해서 적절한 분석이 불가능했다고 판단된다. [그림 8]¹²⁾는 적용한 영상과 분석결과를 보여준다.



<그림 7> ‘Raw Run’ analysis result



<그림 8> <Fantasia2000> ‘sorcerer's apprentice’ analysis result

III. 결론

최근 영상제작 기술의 발전에 힘입어 다양한 형태의 영상 콘텐츠가 대량으로 생산되고 있다. 또한 개인 제작자들도 수준 높은 영상물들을 만들어 내는 것이 가능해지면서 단순히 영상 콘텐츠를 재현할 수 있는 기술 구현보다는 소비자가 선호하는 감성 콘텐츠 영상의 재현이 필요한 시기라는 공감대가 형성되어가고 있다.

12) <www.youtube.com/watch?v=bmpt-s-zYLI&t=1919s>, <검색일 2021년11월30일>

그러나 폭발적으로 늘어난 영상제작 기회에 비해 영상 결과물의 완성도에 큰 영향을 미치는 적절한 배경음악의 선정에는 많은 어려움이 존재하고 있다. 고비용이 소요되는 음악 전문가와 협업이 어려운 개인 제작자들의 경우 영상과 어울리지 않은 OST를 삽입하여 전체적인 영상 수준의 저하는 물론 영상 콘텐츠를 통해 전달하려고 하는 메시지가 희석되는 경우가 빈번하게 발생하고 있는 것이 현실이다.

OST는 영상콘텐츠의 문맥과 비례하여 기본적인 정보와 추가적인 정보를 제공하는 기능을 가지고 있다. 즉, 분위기를 조성하고 미학적 에너지를 추가하는 기본적인 기능, 그리고 커뮤니케이션을 강화하고 영상 방향력의 영역에 리듬구조를 보충하는 추가적인 기능을 수행할 수 있다.

본 연구에서는 제작된 영상이 가지고 있는 Scene의 구성요소들(Scene Design Factor - 시간, 동작, 빈도, 색상, 음향, 배경, 조명, 인물의 감정상태 등) 각각의 특징을 자동으로 분석하여 각 파라미터들과 조화를 이루는 음악의 구성요소들(선율, 리듬, 화성, 형식, 빠르기, 음색 등)을 가진 최적의 OST를 자동으로 선정, 영상 제작자에게 추천할 수 있는 지능적 배경음악 추천 프로그램을 개발하고자 하였다. 본 연구를 통해 창출된 결과는 소규모 영상콘텐츠 사업자 및 개인형태의 콘텐츠공급자(Content Provider)들에게 도움이 되어 영상콘텐츠 산업의 전반적인 질적 수준이 높아지는데 기여를 할 수 있을 것으로 기대한다.

참고 문헌

<논문>

- Hwang, Jin Soo , Har, Dong-Hwan, “A Study on the Development of Intelligent OST Matching Automation Program based on the Correlation between the Components of Image and Music”, KOREA SCIENCE & ART FORUM, 2020.
- Jeong, Yoojeong., “A foundational study for developing colour-music image scale”, Graduate School of Hongik University-Jeong, 2014.
- 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.
- Ministry of Education ,Career Education Policy Division ,“Announcement of 2019 Elementary and Secondary Career Education Survey”, 2019.
- Maeng, Chae Jung , Har, Dong-Hwan, “A Study on the Extraction of Image Configuration Parameters for Intelligent OST Matching”, KOREA SCIENCE & ART FORUM, 2019.
- Owen Craigie Meyers, “A Mood-Based Music Classification and Exploration System”, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2007.
- Russell J.A, “Circumplex Model of Affect” Canada, Journal of personality and social psychology University of British Columbia, Vancouver, 2010.
- Schubert.E, “Measuring emotion continuously : Validityand Reliability of the two - dimensional emotion - space”, Australian journal of Psychology, 1999.
- Zheng Jiang, “Automatic Analysis of Music in Standard MIDI Files” Music and Technology Carnegie Mellon Univ., May 2019.

투고일자: 2021년 10월 19일

심사일자: 2021년 11월 2일

게재확정일자: 2021년 11월 22일

A Study on the Development of Intelligent OST Matching Program Based on Image Composition Parameters

Cho, Do Hee

Researcher, Graduate School of Advanced Imaging Science ,Chung-Ang University

Har, Dong-Hwan

Professor, College of Art&Technology, Chung-Ang University

In this study, an 'intelligent OST matching program' was developed to find background music using image recognition methods. To make this possible, convergence content was created by selecting scene analysis elements as 'emotion,' 'color,' and 'pixel motion speed,' and music components as 'pitch,' 'tempo,' and 'dynamics.' In the program, a 'BGM library' was created in advance and various music was classified. In addition, it was possible to select music that matches as much as possible by analyzing the input video and comparing the components with the analysis results of the music stored in the 'BGM library'. In this process, a radial pentagonal graph was used to make it easier to see the comparison process between video and background music. Finally, it was possible to select 'BGM Choice Sensitivity' and 'The Ratio of Background Music to Video Sound,' and if a matching background music could not be found, music selected in advance could be used. I hope that the program produced in this way can be appropriately used for the creative activities of individual video producers.

Key Words : Scene analysis, Image recognition, Music components,
OST matching program, Convergence content.