ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 20-01-35
254

j.inst.Korean.electr.electron.eng.Vol.24,No.1,254~259,March 2020
http://dx.doi.org/10.7471/ikeee.2020.24.1.254

의도적인 공감각 기반 영상-음악 변환 시스템 구현

Implementation of the System Converting Image into Music Signals based on Intentional Synesthesia

배 명 진*, 김 성 일**★

Myung-Jin Bae*, Sung-Ill Kim**★

Abstract

This paper is the implementation of the conversion system from image to music based on intentional synesthesia. The input image based on color, texture, and shape was converted into melodies, harmonies and rhythms of music, respectively. Depending on the histogram of colors, the melody can be selected and obtained probabilistically to form the melody. The texture in the image expressed harmony and minor key with 7 characteristics of GLCM, a statistical texture feature extraction method. Finally, the shape of the image was extracted from the edge image, and using Hough Transform, a frequency component analysis, the line components were detected to produce music by selecting the rhythm according to the distribution of angles.

요 약

본 논문은 사전에 학습된 기억으로 공감각 현상을 지각할 수 있는 의도적인 공감각으로 영상에서 음악으로 변환하는 시스템을 구현하였다. 영상에서 변환정보로 색상(Color), 질감(Texture), 모양(Shape)을 사용하여 음악의 멜로디(Melody), 하모니(Harmony), 리듬(Rhythm) 정보로 변환하였다. 정적인 영상에서 단조로운 음이 반복되는 것을 최소화하고 영상에 있는 정보를 표현하기 위해 색상의 분포도에 따라 확률적으로 멜로디를 선택하여 출력함으로써 자연스럽게 음을 구성할 수 있도록 하였고, 영상에서 질감은 통계적 질감 특징 추출방식인 GLCM(Gray-Level Co-occurrence Matrix)의 7가지 특징으로 하모니의 장조와 단조를 표현하였다. 마지막으로 모양은 영상의 외곽선을 추출한 후 주파수 성분 분석인 허프 변환(Hough Transform)을 이용해 선 성분을 검출하여 각도의 분포에 따라 리듬을 선택하는 방식으로 음악을 생성하였다.

Key words : Visualization, Sonification, Sound, Music, Color, Texture, Shape, Melody, Harmony, Rhythm

Ⅰ. 서론

과학 기술의 발전으로 인간은 복잡한 정보를 쉽고 빠르게 인지하기 위해 대부분 정보를 시각화하여 사용하고 있다. 이는 인간이 시각정보를 다른 감각보다 빠르게 처리할 수 있고 쉽게 인지할 수 있는 시각적 동물이기 때문이다. 이렇게 시각화하는 과정에서 감각의 전환이 이루어지는데 예를 들

* Dept. of Convergence IT Engineering, Kyungnam University
** Dept. of Electronic Engineering, Kyungnam University
★ Corresponding author
E-mail : kimstar@kyungnam.ac.kr Tel : +82-55-249-2632

※ Acknowledgment
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03932688)
Manuscript received Mar. 6, 2020; revised Mar. 21, 2020; accepted Mar. 27, 2020.

면 소리를 파형으로 나타내 본다거나 적외선 이미지를 이용해 열을 시각으로 전환하는 것이다.

감각의 전환이 자연스럽게 이루어지는 현상을 공감각(Synesthesia)이라 한다. 공감각은 뇌의 정보를 전달하는 부위인 시상(Thalamus)에서 정보를 전달하는 과정에서 일어나며 정보의 혼선으로 일어나는 현상으로 볼 수 있다[1]. 이러한 현상을 통해 일어나는 공감각은 많이 사용하는 감각인 시각에서 다른 감각으로의 변환이 많이 이루어진다. 대표적인 공감각은 색에서 소리가 들리는 색청 공감각이고 문자에서 색이 보이는 자소-색 공감각 등이 있다. 감각의 전달은 기억으로도 이어지며 후각 또는 미각을 통해 특정 기억이 떠오르는 것으로 뇌의 정보전달 혼선 현상은 감각간의 전달뿐 아니라 기억정보에서도 이루어지는 것을 알 수 있다. 인간이 인지할 수 있는 감각은 5가지 감각으로 시각, 청각, 후각, 미각, 촉각으로 나눌 수 있지만, 감각의 인지적 측면에서 보면 근육을 움직이는 정도를 알 수 있는 근감각과 열을 느낄 수 있는 열감도 감각으로 생각할 수 있다. 인간이 7가지 감각을 느낄 수 있다고 보고 이러한 감각들이 공감각을 일으킨다면 총 42가지 공감각을 인지할 수 있으며, 그중 20여 가지 감각의 전환은 현대 기술로 구현할 수 있다고 보고 있다. 공감각으로 감각의 전이를 느끼게 되면 한 가지 감각에서 느낄 수 있는 감각보다 더 많은 정보를 받아들일 수 있다. 선천적인 공감각 외에도 일반인이 학습을 통해 감각 전이 현상을 느낄 수 있는 방법을 의도적인 공감각(Intentional Synesthesia)이라고 한다[2].

의도적인 공감각을 통해 공감각을 일으키면 다양한 정보를 동시에 인지할 수 있다는 장점이 있다. 예를 들면 몇 가지 시각정보를 동시에 인지하는 것은 한계가 있지만 시각정보를 청각정보로 변환하여 변환된 방법을 학습하고 있다면 시각정보를 처리하면서 청각정보를 인지하여 처리할 수 있음으로 다중인지가 가능하다. 의도적인 공감각에서 변환 규칙을 정해놓는다면 규칙에 따라 변환된 결과는 유추가 가능함으로 학습이 필요하지만 동시에 여러 가지 작업을 할 수 있다는 장점은 일의 능률을 올릴 수 있을 것이다. 이러한 과정을 통해 감각이 확장되어 유용한 새로운 기술을 생산할 수 있을 것이다.

본 논문에서는 의도적인 공감각의 장점을 이용해 시각정보인 영상을 청각정보인 음악으로 변환하여 자연스럽게 음악을 들으면 영상정보를 유추할 수 있도록 구현하였다. 청각정보인 음악의 경우 시각정보인 영상에서처럼 약간의 변화를 정밀하게 인지하기는 쉽지 않으며 사람마다 정보를 인지하는 정도에 편차가 있음으로 변환 시 정보의 손실은 감수해야 한다. 하지만 민감도가 낮은 정보를 자연스럽게 전달하는 방법으로 사용할 수 있다.

본 논문에서는 의도적인 공감각을 기반으로 영상에서 음악으로의 변환 방법을 제시하며, 영상에서 변환정보로 색상(Color), 질감(Texture), 모양(Shape)을 사용하여 음악의 멜로디(Melody), 하모니(Harmony), 리듬(Rhythm) 정보로 변환하였다. 정적인 영상에서 단조로운 음이 반복되는 것을 최소화하고 영상에 있는 정보를 표현하기 위해 색상의 분포도에 따라 확률적으로 멜로디를 선택하여 출력함으로써 자연스럽게 음을 구성할 수 있도록 하였고, 영상에서 질감은 통계적 질감 특징 추출방식인 GLCM(Gray-Level Co-occurrence Matrix)의 7가지 특징으로 하모니의 장조와 단조를 표현하였다. 마지막으로 모양은 영상의 외곽선을 추출하여 주파수 성분 분석인 허프 변환(Hough Transform)을 이용해 선 성분을 검출하여 각도의 분포에 따라 리듬을 선택하는 방식으로 음악을 생성하였다.

의도적인 공감각을 이용한 영상에서 음악으로 변환하는 연구가 활성화된다면 시각 장애인을 위한 편의 장비의 개발 또는 영상에 대한 음악적 표현 방법 개발, 공감각 기반 로봇 시스템 및 컴퓨터 인터페이스 등 새로운 감각의 인지 방식이 광범위하고 다양한 응용 분야에서 활용될 수 있을 것으로 기대한다.

II. 관련 연구

공감각을 느끼는 공감각자에 대한 연구는 다양한 방법으로 진행되었으며 23명 중 1명에게 공감각이 일어난다는 결과도 있고 유전적 요인 또는 선천적 요인이라는 연구도 존재한다[3][4]. 하지만 사람마다 느끼는 공감각이 다르고 공감각의 유발 요인이 다르므로 원인을 특정할 수 없다. 본 논문에서는 선천적 또는 공감각자에 대한 연구보다는 비공감각자의 감각의 전환, 학습을 통한 의도적인 공감각 연구에 한정한다.

공감각을 이용한 연구는 시각과 관련된 연구가 많으며 실제 공감각자의 비율도 다른 감각보다 시각 공감각자가 많음으로 시각정보에 대한 연구가 많이 진행되었다. 색과 음의 사이에서 공감각적 변환은 색청 공감각으로 색과 음의 과학적 상관관계를 시작으로 연구가 진행되었다. 가시광선의 파장 대역 중 380~780nm에서 나타나는 색에서 빨간색(650nm), 초록색(520nm), 파란색(433nm)의 파장 비율이 1 : 4/5 : 2/3로 일정한 비율로 나타난다는 것과 가청 주파수 대역인 20Hz~20kHz에서 도, 미, 솔의 음계 파장 비율은 1 : 4/5 : 2/3로 가청 주파수 대역과 가시광선의 파장 대역 비율이 같다. 동일한 파장 대역 비율을 갖는 현상을 기반으로 색의 3요소(색상, 명도, 채도)와 음의 3요소(음계, 옥타브, 음의 세기)를 비슷한 속성으로 간주하여 상호 변환한다[5][6].

공감각 학습의 효과에 대한 연구는 공감각자가 아닌 사람들을 대상으로 9주 동안 공감각 학습 훈련을 실행한 뒤 지능 검사 및 공감각 검사를 진행한 결과, 공감각 학습 훈련을 실행한 사람에 비해 훈련하지 않은 사람이 지능지수(IQ) 평균이 약 12점 더 낮게 나왔고, 학습을 진행한 사람은 공감각자와 비슷한 경험을 할 수 있다고 한다. 문자-색상 공감각자에 대한 연구 중 문자에서 색상을 느끼는 자소-색 인지 패턴이 비슷한 그룹을 대상으로 연구를 진행한 결과, 자소-색 공감각자들이 유아기 시절에 가지고 놀던 자소 장난감과 관련이 있다는 사실을 발견하였다. 자소 장난감은 알파벳 모양의 블록으로 자소에 특정한 색이 있는 장난감이다. 실험에 참여한 공감각자들은 비슷한 자소-색 패턴을 나타내었고, 이 패턴이 유아기 때 가지고 놀던 자소 장난감과 대부분 일치했다. 동일한 장난감을 사용한 공감각자들의 자소-색 공감각 패턴이 유사하다는 연구 결과는 유아기 시절 본인도 모르게 진행된 공감각 학습을 통해 공감각 능력이 발달했다고 생각해 볼 수 있다. 이러한 연구를 통해 공감각은 의도적으로 만들어낸 특정 규칙을 학습하면 타고난 공감각자가 아니더라도 공감각 전이 현상을 느낄 수 있을 것이다[7][8].

본 연구자는 이전 연구에서 근육의 이완 및 긴장 상태를 느낄 수 있는 감각인 근감각을 이용하여 색과 음으로 변환하는 연구를 진행하였다. 이 연구는 앞에서 설명한 색과 음의 유사성에 근감각을 추가하여 움직임을 색과 음으로 변환해 시각 또는 청각만을 이용해 움직임을 유추할 수 있는 연구이다[9]. 해당 연구와 같이 앞으로는 감각과 감각간의 연결이 1:1을 벗어나 1:N의 감각으로 전달이 가능하면 다양한 정보를 동시에 인지할 수 있는 능력을 키울 수 있을 것이다.

III. 제안하는 변환 시스템

영상정보가 음악으로 변환되어 출력으로 나오는 과정을 그림 4에 나타내었다. 음악정보는 쉽게 접근하여 확인할 수 있도록 웹 브라우저에서 출력되도록 구성하였으며, 영상정보는 RGB 카메라를 통해 입력되고 입력된 영상은 임베디드 장비로 보내진다. 웹에서 영상정보를 요청하면 임베디드 보드로 보내진 영상정보는 컴퓨터 비전 라이브러리인 OPENCV를 이용해 영상의 요소인 색상, 질감, 모양으로 나누어 처리된다.

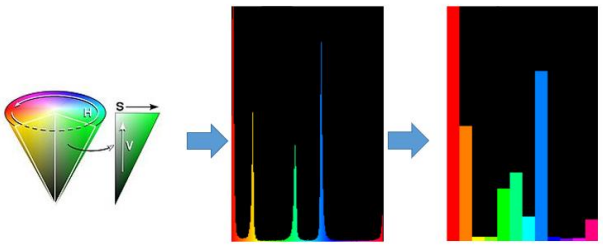

Fig. 1. Color Normalization, (a) HSV Color Model, (b) HSI 0~360° Color Dstribution, (c) 12Color Normalization.
그림 1. 색 정규화, (a) HSI 컬러모델, (b) HSV 0~360° 색분포, (c) 12색상 정규화

색상은 그림 1과 같이 RGB컬러 모델을 HSV 컬러모델로 변환하여 0~360° 색상(Hue)정보를 추출한다. 이때 30° 간격으로 색상 분포도를 만들어 12가지 색상 분포를 추출한다.

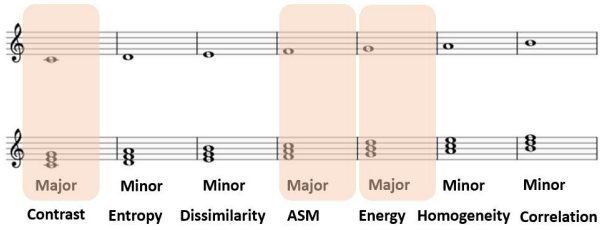

Fig. 2. GLCM Features, Major and Minor.
그림 2. GLCM 특징 및 장조와 단조

질감의 경우 질감특징 검색에서 사용하는 GLCM의 7가지 특징정보(Contrast, Entropy, Dissimilarity, ASM, Energy, Homogeneity, Correlation)를 파이썬 영상처리 라이브러리인 Scikit-image를 이용해 계산 후 각 값을 0~1 사이의 값을 가지도록 정규화한다.

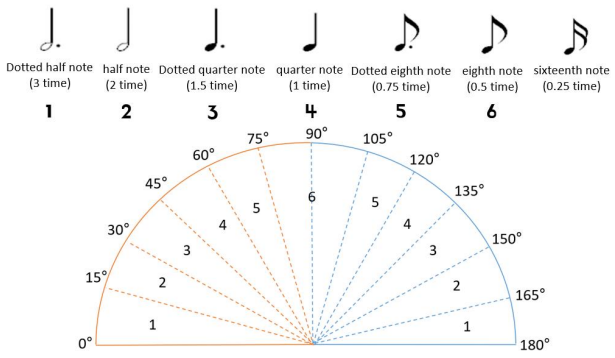

Fig. 3. Hough Transform Normalization and 6 Rhythm.
그림 3. 허프 변환 정규화와 6개의 리듬

그리고 모양의 경우 영상에 캐니 엣지(Canny Edge)를 통해 외곽선 검출을 하고, 허프 변환(Hough Transform)을 이용해 외곽선의 각도에 따른 기울기 분포를 추출한다. 기울기의 경우 0~90°로 표시하며 90~180°는 방향만 다른 직선이므로 0~90° 각도에 맞추어 15° 간격으로 총 6개의 값이 나오도록 분포도에 추가한다. 이렇게 변환된 영상정보들은 node.js 웹서버로 전송되어 저장된다.

저장된 정보들은 웹 요청시 웹 브라우저로 전송되며 브라우저에서는 수신된 값을 음악으로 변환하고 출력 대기한다. 웹 브라우저에서 소리의 출력은 소리 생성 및 출력 라이브러리인 Tone.js를 사용하며 MIDI(Musical Instrument Digital Interface) 신호로 변환하여 출력한다.

음악의 각 요소는 음계, 하모니, 리듬으로 기본적으로 모든 마디는 4/4 박자에 맞추어 출력한다. 음계는 12가지 색상정보 분포도에 따라 랜덤으로 선택하며, 하모니는 질감 특징 정보인 GLCM의 7가지 특징으로 장조와 단조를 선택하여 출력한다. 마지막으로 리듬은 허프 변환으로 얻은 6개의 기울기 값으로 0°에 가까울수록 인정적인 점2분음표(3박) 긴 음을 출력하고, 90°에 가까울수록 짧은 8분음표(0.5박)를 출력한다. 만약 점8분음표(0.75박)를 출력하게 되었을 때 4박에 맞춰 마디를 마무리할 수 없는 경우 16분음표(0.25박)를 넣어서 4박에 맞춘다.

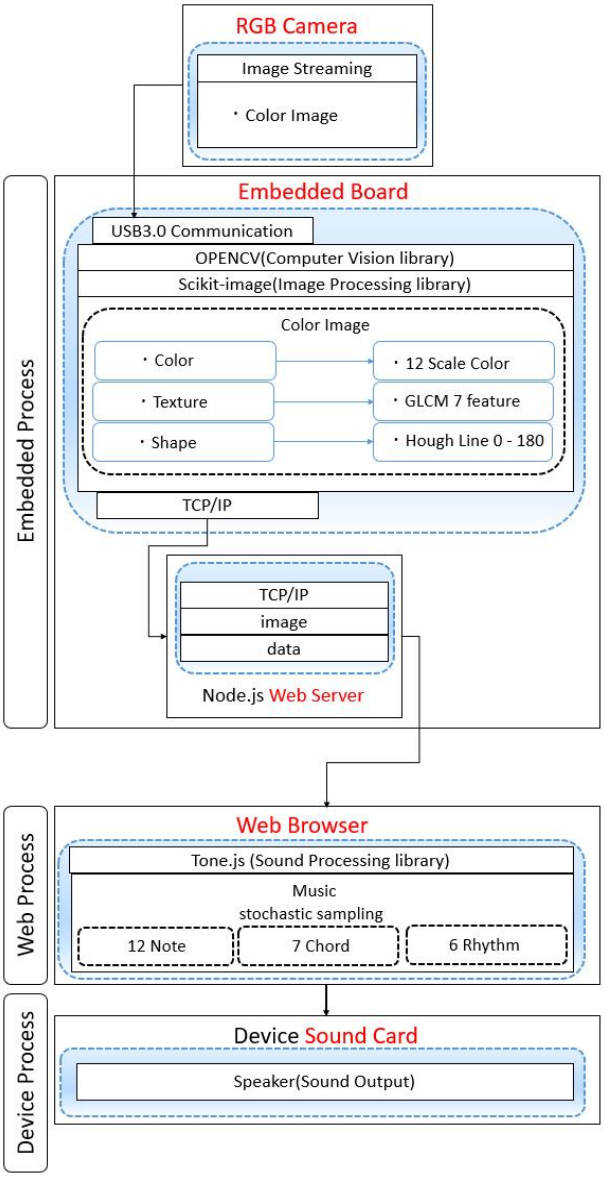

Fig. 4. System configuration.
그림 4. 시스템 구성도

Ⅳ. 실험 및 분석

본 논문의 실험은 제안한 변환 시스템에 2개의 입력 영상을 넣어서 출력 결과를 비교해 보는 것으로 변환 결과를 분석하여 제안한 방법에 맞게 변환되어졌는지 확인하는 것이다. 본 실험에서 사용한 임베디드 디바이스는 엔비디아(Nvidia)사에서 출시한 Jetson Nano 보드다. Jetson Nano 보드에 인텔(Intel)사의 RealSense D415 RGB카메라를 사용한다. 입력 영상은 그림 5의 (a), (d)를 입력하였고 색상 분포도를 구한 값은 (c), (f)와 같다. 색상 분포도를 보았을 때 (d)의 영상이 초록, 파랑 값이 많아 중

간영역에 값이 집중되어 있는 것을 확인할 수 있다.

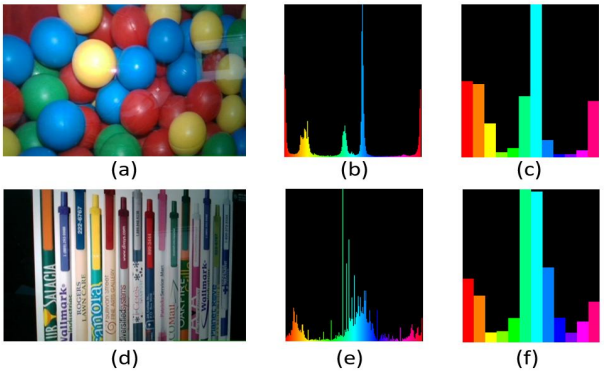

Fig. 5. Input Image and Color, (a) Input Image 1, (b) HSI 0~360° Color Dstribution, (c) 12Color Normalization (d) Input Image 1, (e) HSI 0~360° Color Dstribution, (f) 12Color Normalization.
그림 5. 입력 영상과 색상, (a) 입력영상 1, (b) HSV 0~360° 색분포, (c) 12색상 정규화 (d) 입력영상 2, (e) HSV 0~360° 색분포, (f) 12색상 정규화

그림 6은 입력영상을 허프변환하여 각도 값을 추출한 것이다. 입력 영상 1과 2는 둥근 이미지와 직선 이미지로 결과 값만 보았을 때 결과가 바뀐 것으로 보이지만 입력 영상 2는 짧은 직선 성분들이 많아 값이 추출되지 못하였다.

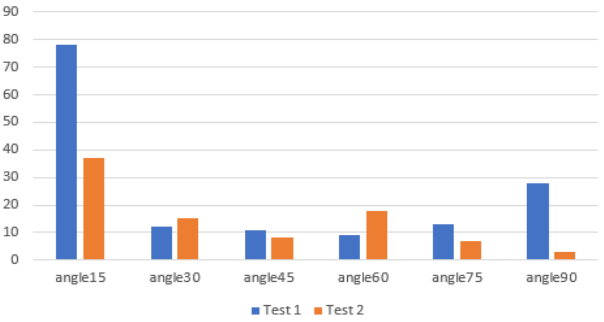

Fig. 6. Input Image Hough Transform Normalization Results.
그림 6. 입력 영상 허프 변환 정규화 결과

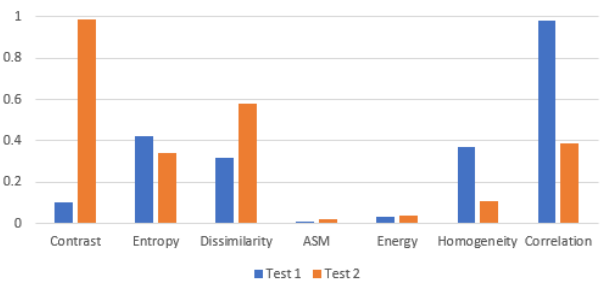

Fig. 7. Input Images GLCM Features.
그림 7. 입력 영상 GLCM 특징

그림 7은 GLCM 특징을 나타낸 결과이고, 이 값으로 하모니를 결정하게 된다. Contrast와 Correlation 값의 경우 차이가 많이 나는 것을 볼 수 있다.

그림 8과 9는 입력 영상 1과 2의 음악변환 결과이다. 총 8마디로 구성되어 있으며 4/4박자에 맞게 음이 추출된 것을 확인할 수 있다.

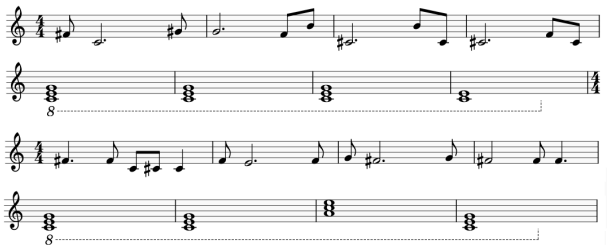

Fig. 8. Music Conversion Results of Input Image 1.
그림 8. 입력 영상 1의 음악변환 결과

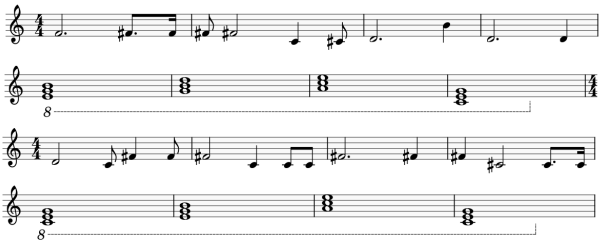

Fig. 9. Music Conversion Results of Input Image 2.
그림 9. 입력 영상 2의 음악변환 결과

V. 결론

본 논문에서는 학습으로 공감각 정보를 얻을 수 있는 의도적인 공감각을 기반으로 영상에서 음악으로 변환하는 방법을 제안하였고 임베디드시스템으로 구현하였다. 구현 결과, 추출한 색상, 질감, 모양의 특징에 따라 음악의 구성이 다르게 출력되는 것을 확인하였다. 하지만 추출한 특징의 분포도에 따라 확률적인 방식으로 값을 추출하기 때문에 실제 출력된 결과가 기대 결과와 일치하지는 않는다. 하지만 실험에서 8마디보다 긴 시간 동안 음악 성분을 추출한다면 결과적으로는 분포도에 맞게 추출이 될 것이다. 본 연구에서는 영상과 음악의 대표적인 특징을 추출하여 변환하는 방식을 채택하였다. 향후 연구에서는 보다 다양한 감각 요소들을 영상과 음악에 적용할 예정이다.

References

[1] J. Ward, "Synesthesia," *Annual Review of Psychology,* Vol.64, No.1, pp.49-75, 2013.
DOI: 10.1146/annurev-psych-113011-143840
[2] Suslick, S Kenneth. "Synesthesia in science and technology: more than making the unseen visible," *Current opinion in chemical biology,* Vol.16, No.5, pp.557-563, 2012.
DOI: 10.1016/j.cbpa.2012.10.030
[3] Smilek, D., Dixon, M. J. & Merikle, P. M. "Synaesthesia: discordant male monozygotic twins," *Neurocase,* Vol.11, No.5, pp.363-370, 2005.
DOI: 10.1080/13554790500205413
[4] Smilek, D. et al. "Synaesthesia: a case study of discordant monozygotic twins," *Neurocase,* Vol.8, No.4, pp.338-342, 2002.
DOI: 10.1076/neur.8.3.338.16194
[5] S. Kim, "Conversion of Image into Sound Based on HSI Histogram," *The Journal of the Acoustical Society of Korea*, Vol.30, No.3, pp.142-148, 2011.
DOI: 10.7776/ASK.2011.30.3.142
[6] M. Bae, S. Kim, "Implementation of the Visualization and Sonification System of Human Movements based on Intentional Synesthesia," *Journal of Korean institute of intelligent systems*, Vol.28, No.1, pp.83-90, 2018.
DOI: 10.7471/ikeee.2018.22.2.362
[7] D. Bor, N. Rothen, D. Schwartzman, S. Clayton, A. Seth, "Adults Can Be Trained to Acquire Synesthetic Experiences," *Scientific Reports 4*, No.7089, 2014. DOI: 10.1038/srep07089
[8] Witthoft, Nathan, and Jonathan Winawer. "Learning, memory, and synesthesia," *Psychological science*, Vol.24, No.3, pp.258-265, 2013.
DOI: 10.1177/0956797612452573
[9] M. Bae, S. Kim, "The System of Converting Muscular Sense Into both Color and Sound Based on the Synesthetic Perception," *Journal of Korean institute of intelligent systems*, Vol.24 No.5, pp.462-469, 2014. DOI: 10.5391/JKIIS.2014.24.5.462

BIOGRAPHY

Myung-Jin Bae (Member)

2013 : BS degree in Electronic Engineering, Kyungnam University.
2016 : MS degree in Conversions IT Engineering, Kyungnam University.
2016~ : PhD course in Conversions IT Engineering, Kyungnam University.

Sung-Ill Kim (Member)

1994 : BS degree in Electronic Engineering, Yeungnam University, Korea
1997 : MS degree in Electronic Engineering, Yeungnam University, Korea
2000 : PhD degree in Computer Science & Systems Engineering, Miyazaki University, Japan
2000~2001 : Research Engineer, National Institute for Longevity Sciences, Japan
2001~2003 : Research Engineer, Center of Speech Technology, Tsinghua University, China
2003~ : Professor, Electronic Engineering, Kyungnam University, Korea