

시·청각 인지능력 향상을 위한 공감각 기반 입체영상-입체음향 변환 시스템

배명진¹·김성일^{2*}¹경남대학교 첨단공학과 박사과정^{2*}경남대학교 전자공학과 교수

The System Converting 3D-Image into 3D-Sound based on Synesthesia for improving Audiovisual Cognitive Ability

Myung-Jin Bae¹ · Sung-Il Kim^{2*}¹Doctor's Course, Department of Convergence IT Engineering, Kyungnam University, Changwon, Korea^{2*}Professor, Department of Electronic Engineering, Kyungnam University, Changwon, Korea

[요약]

기술의 발전과 더불어 많은 정보를 전달받기 위한 인지 능력 향상에 대한 요구가 증가하고 있다. 인간의 신경으로 전달되는 감각은 사람에 따라 또 다른 감각으로 전이되는 현상인 공감각의 형태로 변환되어 전달된다. 본 논문은 인지 능력 향상을 위해 가장 많이 사용하는 감각인 시각에서 청각 정보로 변환하는 시스템을 제안하고 구현하였다. 시각 정보인 영상은 색상, 질감, 모양을 추출하였고 청각 정보인 음향은 멜로디, 하모니, 리듬 정보를 추출하였다. 입체 표현은 영상의 깊이와 음향의 3차원 좌표 공간 정보를 사용하였다. 변환 방법은 색상의 분포도에 따라 확률적으로 멜로디를 선택해 출력하였고, 질감은 GLCM의 특징을 하모니의 장조와 단조로 표현하였다. 마지막으로 모양은 허프 변환으로 선 성분을 검출해 각도의 분포에 따라 리듬을 선택하였다. 변환 과정을 평가하기 위해 이미지를 벽에 투사해 시각적 변환을 확인하였고 임의의 물체를 움직여 입체를 표현하였다. 리얼센스 카메라로 촬영 후 생성된 음악을 분석해보면 제안한 변환 과정에 따라 변환되어 출력되는 것을 확인할 수 있었다.

[Abstract]

This paper proposes and implements a system that converts visual information, which is the most used sense, into auditory information to improve cognitive ability. For visual information of image, color, texture, and shape were extracted, and for auditory information of sound, melody, harmony, and rhythm information were extracted. For the 3D representation, information on the depth of the image and the 3D coordinate space of sound were used. As for the conversion method, a melody was selected and output probabilistically according to the color distribution, and the texture expressed the characteristics of GLCM in major and minor harmonies. Finally, the shape was selected according to the distribution of angles by detecting line components of Hough transform. To evaluate the transformation process, the image was projected on the wall to confirm the visual transformation, and a three-dimensional expression was expressed by moving an arbitrary object. When analyzing music generated by photographing using a realSense camera, it can be confirmed it was converted and output according to the proposed conversion process.

색인어 : 입체영상, 입체음향, 인지능력 향상, 공감각, 3차원 공간정보**Keyword** : 3D-Image, 3D-Sound, Ability Improvement, Synesthesia, 3 Dimensional Information<http://dx.doi.org/10.9728/dcs.2022.23.1.141>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 21 October 2021; Revised 02 November 2021

Accepted 05 January 2022

***Corresponding Author; Sung-Il Kim**

Tel: +82-55-249-2632

E-mail: kimstar@kyungnam.ac.kr

I. 서론

컴퓨터, IT 기기의 소형화와 빠른 통신 속도 등 과학 기술과 정보통신의 발전으로 인간은 복잡한 정보를 빠르게 습득할 수 있게 되었다. 근래에 기술의 융합으로 새로운 기술을 만들어 인간의 인지 능력을 향상시키려는 연구가 활발하게 이루어지고 있다. 단순하게는 웨어러블 디바이스를 이용한 정보 전달 기기에서 시작하여 바이오 해킹을 통한 신체 능력 강화 또는 뇌파 신호를 측정하는 BCI(Brain Computer Interface)를 이용한 BMI(Brain Machine Interface) 기술은 사람의 생각만으로 외부 기기 및 환경을 제어할 수 있다. 사지 마비 환자의 외부 의사소통을 지원하기 위한 기술에서 시작되어 정신질환 진단, 노약자와 장애인들의 재활 및 생활 보조 영역으로 확대되고 있다. BCI를 이용한 BMI 기술은 뇌파를 이용하기 때문에 측정 잡신호 및 개인별 다양한 특성으로 인해 일괄적인 적용보다는 개인 맞춤형으로 사용한다[1].

뇌에 직접 전극을 심는 침습적(invasive) 방법도 일부 환자에게 제한적으로 사용되고 있지만 향후 뇌에 침을 넣는 방식을 사용할 것으로 예상된다. 또한 가상현실(VR: Virtual Reality)과 증강현실(AR: Augmented Reality)은 대표적인 인지 능력 향상 기술이라 할 수 있다. 하지만 인간의 정보 인지 능력에는 한계가 존재한다. 정보 인지에 있어서 시각, 청각, 촉각, 미각, 후각의 5가지 감각에 의존하고 있으며 대부분 정보를 시각화하거나 청각화하여 활용하고 있다. 시·청각 정보를 주로 사용하는 이유는 시각이 다른 감각에 비해 많은 정보를 빠르게 처리할 수 있고 다음으로 청각의 정보 습득이 빠르기 때문이다.

사람에 따라서 하나의 감각으로 다른 감각을 느낄 수 있는 공감각(Synesthesia) 현상을 느끼는 사람들이 있다. 공감각은 시각과 다른 감각의 감각 전이 현상으로 나타나는 경우가 많으며 하나 이상의 감각 손상으로 다른 감각이 발달한 경우에도 공감각이 발생한다. 공감각은 뇌에서 정보를 전달하는 시상(Thalamus)에서 발생하며 정보의 혼선으로 일어나는 현상이라 할 수 있다. 예를 들면 색에서 소리를 인지하는 색·청 공감각이나 글자의 형태에서 색이 보이는 자소·색 공감각과 같이 오감의 영역에서 서로 다른 감각 간의 전이 현상이 발생하는 경우를 말한다. 감각의 전달은 공감각 외에도 기억과도 연결되며 후각 또는 미각으로 특정 기억이 떠오르는 경우를 들 수 있다.

감각의 인지적 측면에서 보았을 때 오감을 제외하고 근육이 움직이는 정도를 인지할 수 있는 근감각과 열을 느낄 수 있는 열감도 감각을 들 수 있다. 이와 같이 인간이 느낄 수 있는 감각이 7가지라고 보고 이러한 감각들이 감각 간의 전이 현상을 일으킨다면 총 42가지 공감각 현상을 인지할 수 있으며, 절반 정도인 20가지 공감각 현상은 현대 기술로 구현할 수 있다. 이러한 공감각을 다양하게 느낄 수 있다면 한 가지 감각을 사용해 느끼는 감각보다 많은 정보를 인지할 수 있다[2].

공감각 현상을 타고난 사람이 아닌 일반인이 공통적인 공감각을 느낄 수 있고 이러한 정보를 표준화할 수 있다면 한 번에 많은 정보를 인지할 수 있을 것이다. 선천적인 공감각자가 아닌

일반인이 반복적인 학습을 통해 감각 전이 현상을 일으키는 것을 의도적인 공감각(Intentional Synesthesia)이라고 한다[3].

3D 깊이(Depth) 인식 기술은 스테레오(Stereo) 방식과 ToF(Time of Flight) 및 Structured Pattern과 같은 방식이 있다. 스테레오 방식은 두 개의 2D 이미지 센서를 결합하여 측정 대상과의 거리를 측정한다. 측정 방식은 두 이미지 센서의 시점 차이에 따른 피사체의 각도 간격을 측정하여 나타낸다. ToF 방식은 광신호가 반사되어 돌아오는 시간 정보를 획득하여 측정한다. 일반적으로 광신호는 스펙트럼의 근적외선을 사용한다. 음향의 전달 방법은 벽이나 천정과 같이 실내의 장애물에 반사, 회절, 분산, 산란 등의 현상을 발생하는 공간 전달계가 존재하고, 사람의 머리 또는 귓바퀴에 의한 반사, 회절, 공진 등의 현상을 유발하는 머리 전달계로 구분한다. 귀로 전달된 소리의 공간감을 지각하는 주요인은 양 귀에 도달하는 시간 차와 레벨 차이 그리고 스펙트럼의 차이이다. 그리고 소리의 공간감을 유발하는 또 다른 요인으로는 시각 효과, 머리의 움직임 또는 소리의 종류에 따른 친숙도 등이 있다[4].

의도적인 공감각을 통해 공감각을 일으키면 약속된 정보로 감각 간의 변화가 일어나므로 객관적인 정보 전달이 가능하며 다양한 정보를 동시에 인지할 수 있다는 장점이 있다. 예를 들면 동시에 전면과 후면의 시각 정보를 인지하는 것은 한계가 있지만, 시각 정보를 청각화할 수 있는 변환 방법을 학습하고 있다면 시각 정보를 처리하면서 청각화된 시각 정보를 인지하여 처리할 수 있으므로 다중인지가 가능하다. 변환 규칙을 정해놓은 의도적인 공감각은 변환된 정보를 변환 규칙을 학습하고 있는 다른 사람에게 전달하여도 같은 결과를 얻을 수 있을 것이다. 학습이라는 과정이 필요하지만 1개 이상의 감각을 다중 인지할 수 있다는 장점은 일의 능률을 올릴 수 있을 것이다. 이러한 감각의 확장은 인간의 인지 능력 향상과 두뇌의 발달로 이어질 수 있어서 현재 인간의 능력에서 한 단계 진화할 수 있을 것으로 기대한다[5].

II. 관련 연구

색과 음은 과학적인 상관관계가 있으며 색과 음의 변환 연구는 이러한 관계를 바탕으로 연구가 진행되었다. 인간이 시각으로 인지할 수 있는 가시광선 파장 대역 중 380~780nm 사이에서 빨간색(650nm), 초록색(520nm), 파란색(433nm)의 파장 비율은 1 : 4/5 : 2/3로 구성되며 인간의 청각으로 인지 가능한 가청주파수 대역인 20Hz~20kHz에서 도, 미, 솔 음계의 파장 비율이 1 : 4/5 : 2/3로 구성되어 가청 주파수와 가시광선의 파장 대역 비율이 동일함을 알 수 있다. 초기 연구에서 색과 음의 이러한 구성 방식의 공통점을 이용해 색의 3요소(색상, 명도, 채도)와 음의 3요소(음계, 옥타브, 음의 세기)를 상호 변환하는 연구를 진행하였다[6].

공감각자가 아니더라도 학습을 통해 공감각을 느낄 수 있다는 것을 증명하기 위해 비 공감각자들에게 9주 동안 공감각 학

습 훈련을 진행하여 공감각 학습 훈련을 하지 않은 그룹에 비해 학습 훈련을 한 그룹이 지능지수 테스트에서 향상된 결과를 얻었고 공감각자와 비슷한 경험을 할 수 있다는 결과가 나왔다. 문자에서 색상을 느낄 수 있는 자색 공감각자들을 연구한 결과 비슷한 그룹으로 나눌 수 있었고 비슷한 그룹의 사람들은 어렸을 때 가지고 놀던 알파벳 모양의 블록의 색상이 공감각자들이 느끼는 자색 공감각과 관계가 있다는 것을 발견하여 공감각이 학습을 통해 지각 능력이 발달되었다고 볼 수 있다[7].

색과 음의 과학적인 상관관계를 기반으로 한 상호 변환 연구와 더불어, 또 다른 관련 연구에서는 학습을 통해 느낄 수 있는 의도적인 공감각을 기반으로 근육의 감각인 근감각에서 시각 및 청각 정보로 변환하는 연구를 진행하였다. 공감각의 학습은 특정 신체 부위의 움직임 변화를 IMU 센서로 롤(Roll), 피치(Pitch), 요(Yaw)의 세 가지 값으로 변환하여 각 요소에 색의 3요소와 음의 3요소를 각각 대응하여 변환하는 형태이다. 단순한 매칭 형태이므로 학습이 쉽고 변환 결과를 쉽게 유추할 수 있다[8].

이전 연구에서는 색과 음과 같이 기초적인 정보를 이용한 변환을 하였다면 본 연구는 이미지와 음악과 같이 정보량이 많은 데이터에서 사람이 인지하기 쉬운 방식의 특징 정보를 매칭하는 방식으로 변환하는 연구를 진행하였다. 이미지에서 음악으로 변환하는 방법은 이미지의 3요소인 색상(Color), 질감(Texture), 모양(Shape)을 사용하여 음악의 3요소인 멜로디(Melody), 하모니(Harmony), 리듬(Rhythm) 정보로 변환하였다.

Ⅲ. 제안하는 시스템

영상에서 음향으로의 변환 방법은 영상과 음악 요소의 공통점을 기반으로 변환 정보로 영상의 3요소인 색상, 질감, 모양을 사용하여 음악의 3요소인 멜로디, 하모니, 리듬 정보로 변환하였다. 영상의 입체감으로 물체와 카메라와의 거리를 표시하는 깊이(Depth) 정보를 사용하였다. 음향의 입체감으로 3차원 좌표인 공간(Space) 정보를 사용하였다.

영상의 입력 변화가 정적일 때 단조로운 소리가 반복되는 것을 최소화하고 영상정보를 출력하기 위해서 색상의 분포도에 따른 확률적 선택 방식으로 멜로디를 출력하여 자연스러운 음으로 구성할 수 있도록 하였고, 질감의 경우 통계적 질감 특징 추출 방식인 GLCM(Gray-Level Co-occurrence Matrix)의 요소인 7가지 특징을 이용해 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 소리를 생성하였다. 그리고 영상의 입체감과 소리의 공간감을 비교해 변환함으로써 실시간으로 움직이는 대상의 위치를 소리의 공간감으로 느낄 수 있도록 하였다.

영상과 음악의 각 요소는 공통되는 부분을 연결해서 학습

시 이질감이 들지 않도록 하였다. 영상의 컬러와 음악의 멜로디는 색과 음의 상호 변환에 대해 색의 3요소와 음의 3요소를 포함하고 있는 값과 유사하게 포함하고 있다. 영상의 컬러는 색채 공간을 뜻하며 색의 3요소인 색상, 채도 및 명도를 의미한다. 영상 신호에서 의미를 지닌 컬러의 특징 요소를 추출하는 방법으로 컬러의 구조적 정보, 컬러 분포 또는 컬러 정보와 공간 정보를 결합한 방법 등으로 추출할 수 있다. 음악의 요소에서 멜로디는 컬러와 같이 음악의 시간 및 공간적인 질서를 나타내는 특징으로 리듬과 음정의 복합체로 음악의 선율을 의미한다.

또한, 영상의 질감은 물체의 표면 정보로 물체 표면의 거친 정도 또는 밝기 변화 정도, 고유 패턴을 의미한다. 영상 신호에서 질감의 특징 요소를 추출하는 방법으로는 통계적 방법과 구조적 방법 또는 스펙트럼분석 방법 등으로 추출할 수 있다. 음악의 하모니는 화성을 의미하고 음악의 공간적인 질서(패턴)를 나타내는 특징을 가지고 있으므로 영상의 질감과 비슷하다.

마지막으로 영상의 모양은 일종의 영역, 경계선 또는 폐곡선을 뜻하며 영상 전체의 윤곽선 정보를 나타낸다. 모양의 변화에 따라 전체적인 분위기가 달라질 수 있다. 모양의 경우 윤곽선 정보의 비교를 위해 윤곽선의 직선 성분을 추출하여 직선 성분의 개수와 각도 값을 이용하여 전체적인 모양 특징을 추출한다. 음악에서 리듬은 음을 배열하는 시간적인 질서를 나타내는 특징으로 음악 전체의 분위기를 결정하는 점이 영상의 모양과 비슷하다.

질감은 영상에서 패턴과 같이 규칙성을 가지는 영상의 요소 중 하나이다. 질감은 반복성을 가지고 있으므로 질감 특징으로 영상에서 특정 영역을 찾거나 인식하는 곳에 사용된다. 질감은 인간의 눈으로는 구분이 쉽지만, 질감을 하나의 값으로 구분하기는 쉽지 않다. 따라서 질감의 특징을 검출하는 방식으로 통계적 질감 추출 방식인 GLCM(Gray-Level Co-occurrence Matrix)을 많이 사용한다. GLCM은 하나의 값으로 질감을 나타내는 것이 아닌 특징을 가지는 여러 요소의 종합적인 수치로 비슷한 질감인지 아닌지를 판단한다. 따라서 하나의 요소라도 다른 값을 표시하는 경우에는 다른 질감을 가지는 것이라 볼 수 있다.

GLCM은 특정 거리 및 특정 각도에 존재하는 한 쌍의 회색(Gray) 레벨에 대한 상대 빈도를 나타내는 매트릭스로 거리 범위는 1에서 이미지의 크기까지이며 측정 각도는 4방향(0°, 45°, 90°, 135°)이다. 다른 각도와 거리에서 생성된 GLCM은 다른 특성값을 나타낸다. 방향 특성이 다른 영상에서 질감 정보를 추출하는 것은 주로 정확한 각도의 선택에 따라 달라진다.

GLCM의 함수는 입력된 영상에서 특정한 값을 가지는 픽셀 쌍이 특정한 공간 관계에서 얼마나 자주 발생하는지 계산하여 GLCM을 생성한다. 이 행렬에서 통계적 측정값을 추출하는 방법을 통해 입력 영상의 질감 특성을 파악한다. GLCM에서는 여러 가지 통계량을 도출할 수 있지만, 대표적인 4가지로 구분한다. 4가지 특징은 대비, 상관, 에너지, 동질성이다. 대비는 명암도 동시 발생 행렬의 국소적 변화를 측정한다.

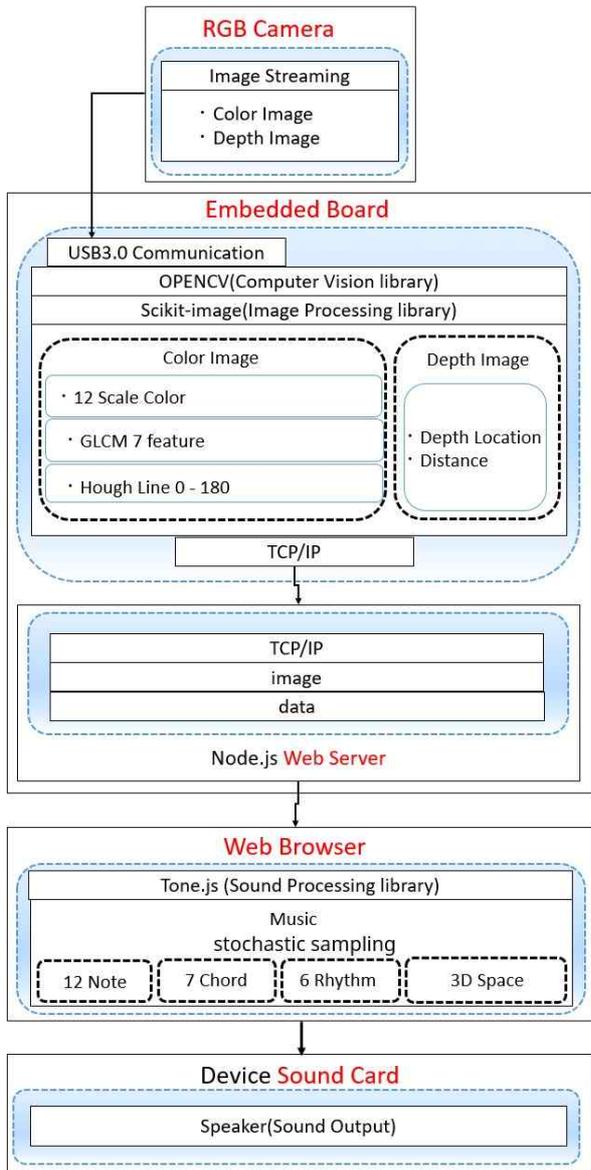


그림 1. 시각-청각 변환 시스템 구조도
Fig. 1. The structure of the image and music conversion system.

상관은 지정된 픽셀 쌍의 결합 확률을 측정하고, 에너지는 GLCM 요소의 제곱 합을 측정한다. 마지막으로 동질성은 GLCM 요소가 GLCM 대각선에 얼마나 근접하여 분포해 있는지 측정한다. GLCM은 이미지에서 인접하는 픽셀의 값의 유무에 따라 GLCM 좌표계를 형성하며 GLCM 좌표계를 정규화하여 값이 1이 넘지 않도록 제한한다. 이렇게 정규화한 GLCM 값을 각 요소에 맞게 수식을 적용하면 질감의 특징을 뽑아낼 수 있다.

사용되는 수식마다 특징을 가지는데 대비(Contrast)와 부동성(Dissimilarity) 그리고 동질성(Homogeneity)은 GLCM 대각선으로부터의 거리와 관련된 가중치를 사용한다. 대비는 픽셀 간 명암 차이를 구분하는 값으로 명암의 분산 정도를 나

타낸다. 인접한 픽셀 간의 대비가 클 때 계산 결과가 크게 나타난다. 따라서 대각선과의 거리가 늘어날수록 가중치가 증가한다. 부동성은 픽셀 간 명암 차이를 구분하는 값으로 픽셀값이 대각선에 모여있지 않으면 큰 값을 나타낸다는 것은 대비와 같지만 가중치가 선형적으로 변한다는 차이점이 있다. 동질성은 행렬 내 각 화소의 균일한 정도를 나타내는 값으로 픽셀값이 대각선에 모여있으면 가장 큰 값을 나타낸다. 다만 균질성은 대비 가중치의 역순으로 값을 가중하며 가중치는 대각선에서 지수적으로 떨어진다.

엔트로피(Entropy)는 픽셀값의 차이가 얼마나 규칙적인지를 나타낸다. ASM은 명암의 균일함을 측정하는 것으로 픽셀 사이의 밝기 변화가 없다면 각 픽셀의 값이 비슷한 값이 되므로 큰 값을 갖는다. 엔트로피는 명암의 임의성을 측정하는 값으로 밝기 값의 변화가 많고 픽셀 값이 랜덤하게 분포되어 있으면 큰 값을 갖는다. 에너지(Energy)는 ASM의 제곱근으로 나타내며 질감 측정 요소로 사용된다. 두 값이 같을 경우는 모든 픽셀이 같을 때 발생한다. 엔트로피는 명암도의 임의성을 측정하는 값으로 픽셀 값의 분포가 퍼져있으면 큰 값을 나타낸다.

마지막으로 연관성(Correlation)은 평균 또는 표준편차와 같은 일반적인 기술 통계량과 유사한 방정식을 사용한다. 연관성은 인접한 픽셀에 대한 선형 종속성을 측정한다. 한 픽셀이 이웃하는 다른 픽셀과 상관되는지에 대한 척도로 값의 범위는 1에서 -1로 표준편차가 0이면 값은 출력되지 않는다.

$$Contrast : \sum_{i,j=0}^{\leq levels-1} P_{i,j} (i-j)^2 \tag{1}$$

$$Dissimilarity : \sum_{i,j=0}^{levels-1} P_{i,j} |i-j| \tag{2}$$

$$Homogeneity : \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{3}$$

$$Energy : \sqrt{ASM} \tag{4}$$

$$ASM : \sum_{i,j=0}^{\leq levels-1} P_{i,j}^2 \tag{5}$$

$$Entropy : H = - \sum_i p_i \log_b p_i \tag{6}$$

$$Correlation : \sum_{i,j=0}^{level-1} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \tag{7}$$

식 (1)-(7)에서 i, j는 양자화 행렬 내의 좌표이고 level은 입력 영상의 밝기 값의 차원으로 밝기 등급의 범위를 나타낸다[9].

영상에서 모양을 특정하기는 쉽지 않다. 단순한 외곽선 추출로 모양을 구분하기에는 구분 지을 수 있는 정보가 부족하다. 외곽선 추출을 이용해 모양을 기술하는 방법으로 푸리에 디스크립터(Fourier Descriptor)와 체인 코드(Chain Code) 같은 방식으로 사용되고 있으며 CSS(Curvature Scale

Space) 방식으로 MPEG-7에서 사용되고 있다. 하지만 특정 모양이 있는 이미지가 아닐 경우 모양을 사람이 특정하여 구분하기도 어렵고 외곽선 검출인 캐니 외곽선 추출 결과를 보면 특정 모양을 검출하여 구분하기 쉽지 않다.

직선 성분을 추출하여 입력된 이미지의 방향성 값인 각도를 측정하여 특정 구역의 각도 값의 누적 분포로 모양 특징을 구분한다. 측정된 직선 성분의 각도 값의 분포에 따라 모양의 특징을 구분할 수 있다. 형태가 불규칙한 이미지의 경우 각도 분포가 불규칙할 것이고 형태가 일정한 패턴을 가지고 있다면 특정 패턴에 맞는 각도 분포를 가질 것이다. 이렇게 직선 성분의 각도 분포로 모양을 검출한다. 허프 직선 검출 알고리즘은 이미지에서 직선을 찾는 영상처리 방식 중 하나로 이미지 내의 어떤 점을 선 집합의 일부로 가정하여 직선 방정식을 이용해 직선을 검출하는 방식이다.

IV. 실험 및 분석

본 논문의 실험은 프로젝터로 벽에 이미지를 출력하고 인텔의 리얼센스(Intel Realsense) 카메라로 해당 이미지를 촬영하는 방식으로 진행하였고 벽과 카메라 사이에 물체를 두어 거리와 위치를 변경하면서 테스트하였다. 카메라를 통해 입력되는 이미지의 색상, 질감, 모양 정보가 청각 정보의 멜로디, 하모니, 리듬으로 변환하는 과정을 확인하고 이미지의 깊이 값이 3차원 공간 정보로 변환되어 설계한 시스템대로 변환이 이루어지는지 확인하였다. 깊이 영상은 1m 안에 들어오는 물체의 제일 가까운 위치 값을 영상의 공간좌표인 x, y 값으로 변환하고 거리값을 z 값으로 변환한다.

<그림2> (a)의 이미지는 산에서 찍은 도시 풍경 사진으로 프로젝터로 정지된 이미지를 출력하여 리얼센스 RGB 카메라로 인식한 결과는 (b)에 표시하였다. 카메라의 위치가 아래쪽에 있어서 아래에서 위쪽을 바라보는 형태로 촬영되었다.

입력된 이미지를 기반으로 세 가지 정보를 추출한다. 첫 번째는 컬러 분포를 이용해 12가지 색으로 정규화하는 과정이고, 두 번째는 7가지 GLCM 질감 특징을 검출하는 과정이다. 마지막으로 세 번째는 허프 라인 검출을 통해 이미지에서 검출된 라인을 15°씩 묶어서 90°를 6 등분하여 구분하는 과정이다.



그림 2. (a) 입력 이미지, (b) 입력된 RGB 이미지
Fig. 2. (a) Input Image, (b) Input RGB Image

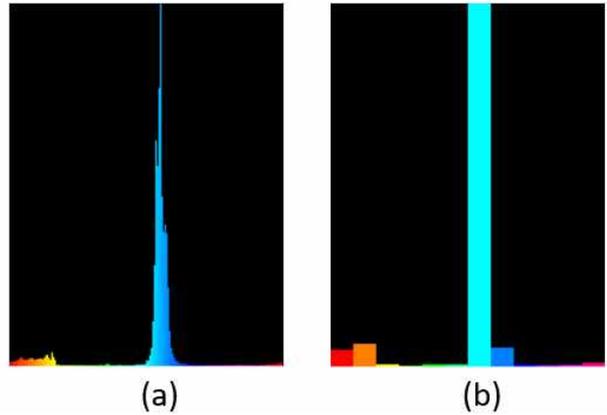


그림 3. (a) 컬러 분포, (b) 12가지 색으로 정규화
Fig. 3. (a) Color distribution, (b) Normalized 12 colors

<그림 3>과 같이 RGB 카메라로 입력된 이미지에서 HSV 컬러 모델에서 색상 분포를 만든 후 12가지 색으로 정규화하여 색상의 분포에 따라 음계를 추출할 수 있도록 하였다.

<그림 4>와 같이 이미지에서 질감의 7가지 특징을 추출하였고, 추출된 값 중 Contrast가 높게 나온 것을 확인하였다. Contrast는 픽셀 간 명암 차이를 구분하는 값으로 명암의 분산 정도를 나타내고 인접한 픽셀 간의 대비가 클 때 계산 결과가 크게 나타나므로 입력된 이미지는 대비가 큰 것을 알 수 있다. GLCM 질감 특징을 분석하였을 때 입력된 이미지의 명암 차이가 뚜렷하게 나타나는 것을 질감 특징값으로 확인할 수 있다.

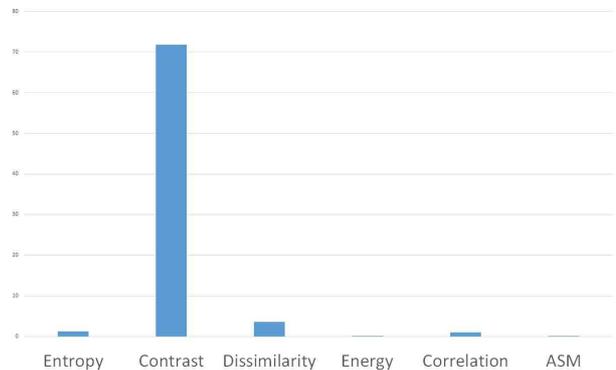


그림 4. GLCM 질감 특징 검출
Fig. 4. Detection of GLCM texture features

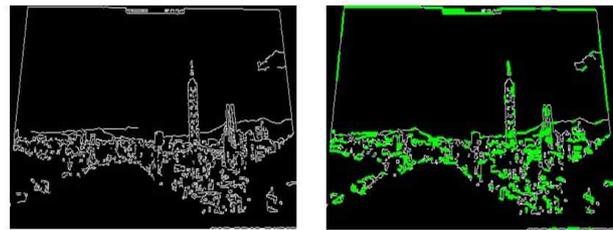


그림 5. (a) 캐니 외곽선 검출, (b) 허프 라인 검출
Fig. 5. (a) Canny edge detection, (b) Hough Line detection

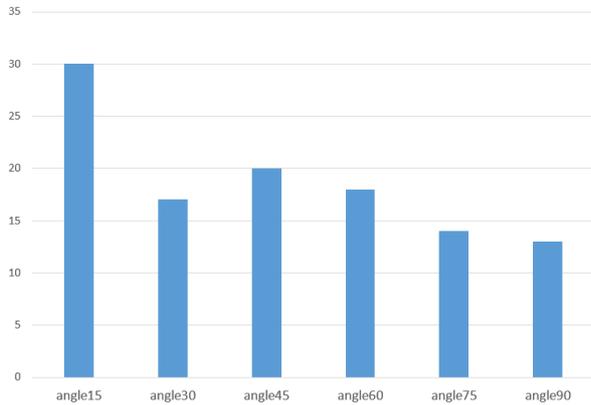


그림 6. 허프 라인 검출을 이용한 직선 성분 각도 추출 결과
 Fig. 6. Results of straight component angle extraction using Hough Line detection.

<그림 5>와 같이 입력 이미지의 모양을 검출하기 위해 캐니 외곽선 검출과 허프 라인 검출을 진행하여 허프 라인 검출에서 검출된 직선의 기울기를 확인하였다. <그림 5>의 (a)는 캐니 외곽선 검출 결과이고 외곽선 추출 결과에서 허프 라인 검출 결과는 (b)에 표시하였다.

<그림 6>은 허프 라인 검출을 이용한 직선 성분의 각도를 추출하고 이를 기반으로 도출한 분포도이다. 여기서 추출된 라인의 각도를 추출한 결과 0°에서 15° 사이의 분포가 가장 많이 추출되었고 나머지 각도는 비슷한 값의 분포로 나타났다.

<그림 7>은 이미지에서 추출한 데이터를 기반으로 음악으로 변환된 결과를 출력하였다. 총 8마디로 되어 있으며 4/4박자로 표시하였다. <그림 3>에서 정규화된 12가지 색상을 확률적으로 추출하여 생성한 음계로 정규화된 데이터를 보편 하나의 색상 분포로 집중된 것을 확인할 수 있으므로 변환된 음계도 해당 음계가 많이 추출된 것을 확인할 수 있다.

질감 특징에 따른 하모니 추출 결과로 1도 화음이 4개로 가장 많이 추출되었고 다음으로 나머지 화음들이 하나씩 들어가 있는 것을 알 수 있다. 즉, 질감의 특징적 확률 분포에 따라 해당 하모니 성분으로 제대로 변환된 것을 확인할 수 있었다.

허프 라인 검출을 이용한 리듬의 변화는 점2분음표가 가장 많이 나타났다. 점2분음표가 3박을 의미하므로 나머지 1박은 8분음표가 들어가거나 4분음표가 들어가서 4/4박자를 채우게 된다.



그림 7. 음계, 하모니, 리듬 결정 결과
 Fig. 7. The Results of the Scale, Harmony and Rhythm



그림 8. (a) RGB 이미지, (b) 깊이 이미지
 Fig. 8. (a) RGB Image, (b) Depth Image

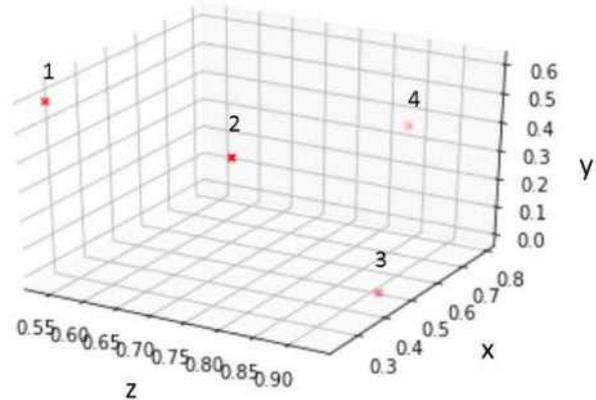


그림 9. 소리의 3차원 공간좌표
 Fig. 9. 3D Space of Sound

네 번째 마디의 경우 처음에 8분음표가 들어갔고 두 번째로 점2분음표가 들어가서 마지막 박자는 8분음표가 채워진 것을 확인할 수 있다.

<그림 8>은 이미지와 리얼센스 카메라 사이에 막대를 두어 촬영된 이미지로 깊이 이미지로 확인해 보면 물체가 감지되는 것을 확인할 수 있다. 이렇게 물체가 감지되면 해당 물체의 가장 가까운 좌표를 인식하게 되고 이미지의 x, y 좌표와 물체와의 거리 값 z축으로 표시할 수 있다. 물체의 위치를 상하좌우로 이동하며 좌표 값을 추출하였고 추출한 값을 소리의 차원 좌표계로 변환하여 출력하므로 소리를 들을 때 소리가 나는 위치가 변경되며 들리게 된다. <그림 9>는 소리로 출력되는 3차원 좌표 값을 표시하였다. 소리의 중앙 지점은 x, y, z 축의 각 값이 0.5인 위치이고 이 위치를 기준으로 상하좌우가 결정된다. 카메라에 검출될 수 있는 물체의 거리는 최대 1m로 고정하였고 1m 이내에 들어오는 값을 z축 좌표에 0~1 사이로 표시하였다. 가까운 값은 0이고 멀리 있으면 1로 표시된다. x축과 y축은 가로와 세로 값을 0~1 사이 값으로 정규화하여 표시하였고 이미지의 해상도는 1920 x 1080이다. 따라서 <그림 9>의 1은 카메라에서 0.5m 떨어진 곳의 우측에 있는 물체 위치를 소리로 변환된 좌표를 나타낸다. 2번은 좌측하단 3번은 중앙 하단에서 가장 먼 곳이고 4번은 중앙 좌측을 나타낸다.

V. 결 론

본 논문에서 인간의 시·청각 인지 능력을 향상하는 방법으로 공감각 기반의 시각(입체영상)에서 청각(입체음향)으로 변환하는 방법을 제안하였다. 입·출력 변환 결과를 확인해 보면 색상, 질감, 모양의 특징을 추출하여 음계, 하모니, 리듬에 각각 대응하여 변환한 결과가 제안한 변환 시스템의 의도대로 도출된 것을 확인할 수 있었다. 이미지에서 깊이 값을 추가하여 3차원 좌표계로 변환한 값을 3차원 소리로 변환하여 출력한 소리를 들어보면 물체의 위치가 빠르게 변하면 쉽게 알기는 어렵고 물체가 천천히 이동하는 경우 구분할 수 있었다. 하지만 전후좌우 한 방향으로만 이동한 경우는 인지하기 쉬웠지만 좌측 상단과 같이 대각선 위치로 이동한 경우는 방향을 쉽게 인지하기 어려웠다. 출력 결과는 2채널 서라운드 출력과 7.1채널 출력 결과를 비교해 보면 7.1채널이 소리를 출력하는 스피커가 방향에 따라 구분되어 있어서 약간의 차이가 발생하였지만 큰 차이는 없었다.

제안하는 시스템의 변환 과정을 인지하고 있으면 변환 관계를 유추할 수 있다. 이와 같은 특징을 이용하면 하나의 감각을 이용해 여러 가지 다른 감각의 신호를 얻을 수 있으므로 인지 능력의 향상을 기대할 수 있을 것이다.

감사의 글

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2020R1F1A1069762)

참고문헌

- [1] W. He, Y. Zhao, H. Tang, C. Sun, W. Fu, "A wireless BCI and BMI system for wearable robots", *Journal of cognitive neuroscience*, Vol. 46, No. 7, pp. 936-946, July 2015. <https://doi.org/10.1109/TSMC.2015.2506618>
- [2] K. S. Suslick, "Synesthesia in science and technology: more than making the unseen visible", *Current opinion in chemical biology*, Vol. 16, No. 5, pp. 557-563, Dec 2012. <https://doi.org/10.1016/j.cbpa.2012.10.030>
- [3] M. Bae, S. Kim, "Implementation of the System Converting Image into Music Signals based on Intentional Synesthesia", *Journal of IKEEE*, Vol. 24, No. 1, pp. 254-259, Mar 2020. <https://doi.org/10.7471/ikeee.2020.24.1.254>
- [4] F. Keyrouz, K. Diepold, S. Keyrouz, "High performance 3D sound localization for surveillance applications", *Proceedings of the ninth international conference on virtual systems and multimedia IEEE*, pp. 563-566, Sept 2007. <https://doi.org/10.1109/AVSS.2007.4425372>
- [5] D. Bor, N. Rothen, D. Schwartzman, S. Clayton, A. Seth, "Adults Can Be Trained to Acquire Synesthetic Experiences", *Scientific Reports*, Vol. 4, No. 7089, pp. 1-8, Nov 2014. <https://doi.org/10.1038/srep07089>
- [6] 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, Apr 2011. <https://doi.org/10.7776/ASK.2011.30.3.142>
- [7] N. Witthoft, J. Winawer, "Learning, memory, and synesthesia", *Psychological science*, Vol. 24, No. 3, pp. 258-265, Jan 2013. <https://doi.org/10.1177/0956797612452573>
- [8] 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, Jun 2018. <https://doi.org/10.7471/ikeee.2018.22.2.362>
- [9] K. Lee, S. Jeon, B. Kwon, "Implementation GLCM/GLDV-based Texture Algorithm and Its Application to High Resolution Imagery Analysis", *Korean Journal of Remote Sensing*, Vol. 21, No. 2, pp. 121-133, Apr 2005.



배명진(Myung-Jin Bae)

2013년 : 경남대학교 전자공학과 공학사
 2016년 : 경남대학교 대학원 첨단공학과 (공학석사)
 2016년~현 재: 경남대학교 대학원
 첨단공학과 박사과정

※ 관심분야 : 멀티미디어신호처리, 임베디드시스템, 공감각 등



김성일(Sung-II Kim)

1994년 : 영남대학교 전자공학과 (공학사)
 1997년 : 영남대학교 대학원 전자공학과 (공학석사)
 2000년 : (일본) Miyazaki University (공학박사)

2000년~2001년: (일본) National Institute for Longevity Sciences, 연구원

2001년~2003년: (중국) Center of Speech Technology, Tsinghua University, 연구원

2014년~2015년: (미국) CSUS 방문연구교수

2003년~현 재: 경남대학교 전자공학과 교수

※ 관심분야 : 멀티미디어신호처리, 임베디드시스템, 공감각 등