

강인한 크로마그램 성분 추출을 통한 커버곡 검색 성능 개선

서진수[†]

Improving Cover Song Search Accuracy by Extracting Salient Chromagram Components

Jin Soo Seo[†]

ABSTRACT

This paper proposes a salient chromagram components extraction method based on the temporal discrete cosine transform of a chromagram block to improve cover song retrieval accuracy. The proposed salient chromagram emphasizes tonal contents of music, which are well-preserved between an original song and its cover version, while reducing the effects of timbre difference. We apply the proposed salient chromagram extraction method as a preprocessing step for the Fourier-transform based cover song matching. Experiments on two cover song datasets confirm that the proposed salient chromagram improves the cover song search accuracy.

Key words: Cover Song, Chromagram, Salient Component, Music Retrieval, Music Similarity

1. 서론

컴퓨터, 스마트폰, AI 스피커 등의 정보처리 기기와 유무선 네트워크 기술의 발달에 따라서 사용자가 원하는 각종 콘텐츠를 빠르고 신뢰성 있게 찾아서 제공해 줄 수 있는 검색 기술의 필요성이 커지고 있다[1-3]. 이와 관련하여 사용자의 요구를 능동적으로 반영하여, 음악을 찾아서 제공하는 것을 가능하게 하는 음악 정보 처리 및 검색 기술이 활발히 연구되고 있다. 음악 검색과 관련하여 핑거프린팅 기술이 관심을 받아왔지만, 핑거프린팅의 경우 원본에서 압축, 잡음처리, AD/DA 변환 등의 신호처리 과정을 통해서 파생된 복사본만을 검색할 수 있다[4]. 콘서트 현장에서 라이브 녹음 또는 리메이크를 통해서 재편집

된 음악을 원곡의 커버곡이라 부르며, 이러한 커버곡은 핑거프린팅 방법으로는 검색할 수 없다. 커버곡 검색은 음악 아카이브 정리, 동영상공유매체를 통한 음원 저작권 침해 방지 등에 활용될 수 있다.

커버곡 검색을 위해서는 서로 다른 음원에 대해서는 차별성을 가지고, 커버곡 간에는 공유되는 음악 신호의 특징을 찾아야 한다. 핑거프린팅 및 유사 음악 검색에서는 신호의 음색 특성을 반영하는 주파수 분석을 통한 특징들이 많이 사용되었으나 음색은 개별 가수 및 악기의 특성과 연관된 것으로 커버곡 검색에는 적합하지 않다. 커버곡이 만들어지는 주요한 경로로는 라이브 음악 녹음과 리메이크를 위한 편집 음악이 있다. 라이브 음악 녹음의 경우 가수, 악기, 연주자의 차이로 인한 음색의 변화가 두드러지며, 편

* Corresponding Author : Jin Soo Seo, Address: (25457) 7 Jukhun-gil, Gangneung, Gangwon-do, Korea, TEL : +82-33-640-2428, FAX : +82-33-646-0740, E-mail : jsseo@gwnu.ac.kr

Receipt date : Mar. 25, 2019, Revision date : May 1, 2019
Approval date : May 14, 2019

[†] Dept. of Electrical Eng., Gangneung-Wonju National University

* This research was supported by 2018 Academic Research Support Program in Gangneung-Wonju National University

집 음악의 경우 음색 변화와 함께 템포 및 리듬, 도입부와 코러스부의 위치 변화 등 좀 더 다양한 변형이 발생할 수 있다. 이러한 다양한 차이에도 불구하고 사람들은 두 곡간의 관계가 커버곡인지 여부를 쉽게 판단할 수 있다. 이는 커버곡들 간에 음들의 시간적 연결을 의미하는 선율(멜로디) 정보가 잘 보존되기 때문이다. 이러한 선율 정보를 잘 나타낼 수 있는 신호 특징으로 크로마그램이 있으며, 커버곡 검색에 널리 사용되고 있다. 크로마그램은 인간 청각이 옥타브 차이가 나는 주파수를 가진 두 음을 유사음으로 인지한다는 음악이론에 기반한다. 옥타브 차이나는 음악의 피치 성분들을 가산하여 음악의 전체 주파수 성분들을 하나의 옥타브 안으로 접어서 표현한 것이다. 크로마그램은 주파수 분석에 이은 옥타브 단위로 합산하는 과정에서 가수의 목소리와 악기에 연관된 음색보다는 음악 자체의 화성적 구조에 보다 밀접하게 연관되게 된다[5]. 크로마그램을 얻는 방법에는 다른 음악 스펙트럼 주파수 분석 방법들과 같이 고정된 짧은 길이(수십~수백ms)의 프레임으로 나누어 각 프레임에서 크로마그램 벡터를 추출할 수도 있고, 음악의 비트와 동기를 맞추어 음악 온셋(onset) 사이의 크로마 값들의 평균을 취하여 크로마그램을 출력할 수도 있다. 고정된 프레임을 사용하는 크로마그램은 템포 변화에 취약하므로, 음악의 크로마그램 간 비교 시에 템포 변화를 고려하여야한다. 반면에 비트동기 크로마그램은 커버곡 생성과정 중에 템포 변화가 발생하더라도 강인성을 가지는 장점이 있으나 비트 추출의 정확도가 전체 커버곡 검색 성능에 큰 영향을 미치게 되는 단점이 있다. 크로마그램 추출에 관한 상세한 설명은 참고문헌 [5]와 [6]에 있다.

커버곡 검색 방법에는 크로마그램 수열 직접 비교 방법과 전곡기반 특징 축약 방법의 크게 두 가지로 나누어진다. 수열 직접 비교 방법[7,8]은 음성 인식과 DNA 수열 비교에 사용되어왔던 dynamic time warping이나 Smith-Waterman 거리 등이 커버곡 검색에 적용되어 상대적으로 검색 성공률 보였으나, 검색에 시간과 계산량이 많이 소요되고 전곡의 크로마그램을 모두 저장해야하므로 저장 공간이 많이 필요한 단점이 있다. 반면에 전곡기반 특징 축약 방법[9,10]은 검색에 용이한 고정된 길이의 크로마그램 전곡 특징을 구하여 검색에 필요한 특징 저장 공간을 줄일 수 있고, 일반적으로 유클리디안 또는 해밍거리

를 통한 DB 인덱싱도 가능한 장점이 있지만 수열 직접 비교에 비해서 검색 성능이 떨어지는 단점이 있다. 본 논문은 전곡기반 특징 축약 방법의 하나인 2D FTM(Fourier Transform Magnitude) 기반 크로마그램 축약 방법[10]의 성능을 개선하기 위해서 전처리 과정을 추가하는 방법을 제안한다. 2D FTM 기반 크로마그램 축약 방법은 커버곡 생성과정에서 음악 키의 변화로 인해서 크로마그램 벡터가 크로마 축으로 이동하는 경우가 자주 발생하므로 푸리에 변환의 성질 중 푸리에 변환 크기값의 이동 불변성을 활용하여 음악 키 변화에 대한 강인성을 얻는 방법이다. 본 논문에서는 2D FTM 기반 크로마그램 축약 방법의 성능을 개선하기 위해서 크로마그램 블락(block)을 시간축 이산 코사인 변환(temporal discrete cosine transform, TDCT)[11]하고 강인한 성분만을 남기는 방법을 제안한다. 커버곡에서도 유지가 되는 강인한 크로마그램 성분을 추출하기 위해서 TDCT 절대값이 큰 성분들을 이용하였다. 두 가지 커버곡 데이터셋에서 실험을 수행하여, 기존 2D FTM 기반 크로마그램 축약 방법에 제안한 전처리 과정을 추가함으로써 커버곡 검색 정확도를 10%이상 개선했음을 확인하였다.

본 논문은 크로마그램 수열 축약에 기반한 커버곡 검색에 관한 연구이다. 2장에서 기존 이차원 푸리에 변환을 이용한 수열 축약 방법을 살펴보고, TDCT를 이용한 강인한 크로마그램 성분 검출 방법을 제안한다. 3장에서 제안한 전처리 과정의 유무에 따른 커버곡 검색 성능을 실험하고 결과를 비교 분석한다.

2. 강인한 크로마그램 성분을 이용한 푸리에 변환 기반 커버곡 검색

특징요약 기반 커버곡 검색기는 Fig. 1과 같이 추출된 프레임 레벨 특징을 모아서 요약하여 전곡단위

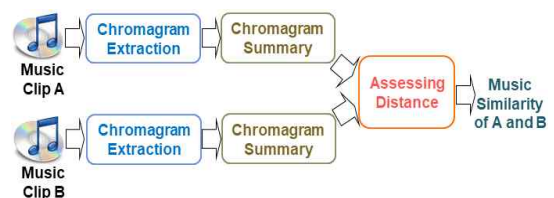


Fig. 1. Music similarity computation based on chromagram summary.

특징을 얻고, 전곡 단위 특징간 거리 비교를 통해서 커버곡 유무를 판정하게 된다. 본 논문은 프레임 레벨 특징 요약 방법으로 2D FTM을 사용하고 거리 비교 방법으로 유클리디안 거리를 사용한 기존 방법 [10]의 성능을 개선하기 위한 전처리 방법에 관한 연구이다. 크로마그램 특징을 2D FTM으로 요약하기 전의 전처리 과정으로 TDCT를 사용하여 강인한 크로마그램 성분을 추출하는 방법을 제안한다. 먼저 2D FTM 기반 특징 요약 방법을 살펴보고, 제안한 전처리 방법인 강인한 크로마그램 성분 추출 방법을 기술한다.

2.1 이차원 푸리에 변환 기반 크로마그램 축약을 통한 전곡 특징 추출

이차원 푸리에 변환을 이용한 크로마그램 축약을 통한 전곡 특징 추출 방법은 Fig. 2와 같다. 음악 신호를 프레임단위로 나누어 크로마그램을 얻는다. 크로마그램은 오디오의 스펙트로그램처럼 고정된 짧은 길이(수십~수백ms)의 프레임으로 오디오 신호를 나누어 각 프레임에서 L 차 크로마그램 벡터(일반적으로 $L=12$)를 추출하게 된다. 고정된 길이의 프레임에서 얻은 크로마그램을 그대로 사용할 경우 커버곡 생성과정에서 조변화와 함께 가장 흔히 발생하는 템포 변화에 대해서 강인성을 가질 수 없다. 2D FTM 방법에서는 템포 변화에 대한 강인성을 얻기 위해서 비트 동기 크로마그램[14]을 사용한다. 비트동기 크로마그램은 크로마그램 추출과 비트 추출을 동시에 수행하고 음악 onset 사이의 크로마그램 값들의 평균을 취하여 출력한 것이다. 음악 신호로부터 얻은

크로마그램을 시간 순으로 수열로 표기하여 $\mathbf{X} = (X_1, X_2, \dots, X_N)$ 이라면, \mathbf{X} 를 L 행 N 열 2차원 신호로 생각할 수 있다. 시간축으로 S 만큼 이동하면서 크로마그램 벡터를 W 개씩 묶어서 크로마그램 블록을 만든다. 각 크로마그램 블록은 L 행 W 열 2차원 신호이며, 크로마그램 블록을 이차원 푸리에 변환한다. 커버곡 생성 과정에서 가장 흔히 발생하는 음악의 조변화는 크로마그램 상에서 크로마축으로 원형이동으로 나타나게 된다. 푸리에 변환의 성질 중에서 신호를 원형이동할 경우 푸리에 변환 상에서 크기는 변화하지 않고 위상만 이동한다는 성질을 활용한다. 즉 신호가 원형이동 되더라도 푸리에 변환의 크기값은 변하지 않으므로, 음악의 조변화에 불변하도록 크로마그램 블록의 푸리에 변환 크기값인 2D FTM을 구한다. 각 블록의 2D FTM들을 전곡 단위로 요약하기 위해서 각 블록의 푸리에 변환 크기값들의 중간값을 구하여 전곡 단위 특징으로 사용한다. 커버곡 검색을 위한 전곡단위 특징 간의 거리는 유클리디안 거리를 사용한다.

2.2 시간축 이산 코사인 변환을 이용한 강인한 크로마그램 블록 추출

본 논문에서는 Fig. 2의 2D FTM 방법에서 크로마그램 블록을 푸리에 변환하기 전에 강인한 성분을 추출하여 성능을 개선한 2D SCFTM (Salient Chromagram Fourier Transform Magnitude) 방법을 제안한다. 크로마그램 각 밴드별로 시간축 방향으로 이산 코사인 변환을 취하고 크기가 큰 성분만을 남기는 방법으로 음악 고유의 시간 방향 음의 강약을 강조하였다. 이러한 과정을 통해서 커버곡들 간에 잘 보존되는 음들의 시간적 연결인 선율 정보 중에서 주선율이 강조되고 부선율과 잡음의 영향은 줄어들게 된다. TDCT를 이용한 시간방향 정보 처리는 음성 인식[11], 비디오 신호 처리[12] 등에 성공적으로 적용되었다.

2.1에서 기술한 2D FTM 방법에서 길이가 W 인 크로마그램 블록을 $f[c, m]$ 이라고 하자. 크로마그램 차수가 L 이라면 $f[c, m]$ 은 L 행 W 열 이산 2차원 신호가 된다. 음악 고유의 시간 방향 음의 강약을 강조하기 위해서 각 크로마그램 밴드별로 시간축 방향으로 길이 W 의 1차원 DCT를 구한다. 2차원 신호 $f[c, m]$ 의 m 번째 크로마그램 밴드($c=m$)에서 시간방향 평균을

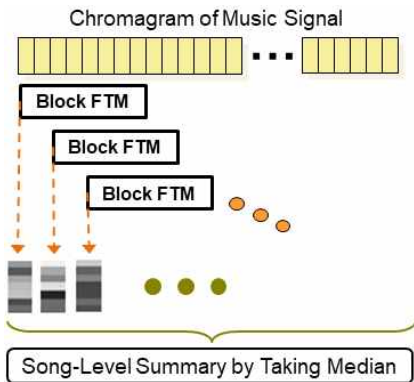


Fig. 2. Summarizing chromagram of a song using 2D FTM [10].

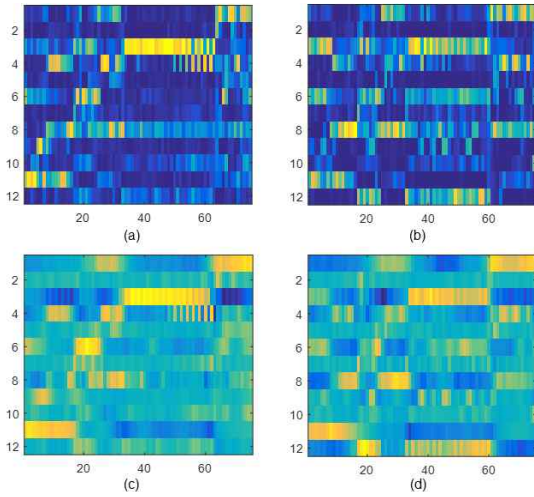


Fig. 3. (a) Chromagram of the excerpt of the original song "I'm losing you". (b) Chromagram of the excerpt of the cover song "I'm losing you". (c) Salient chromagram of (a). (d) Salient chromagram of (b).

차감하고 구한 이산 코사인 변환인 $F_m[k]$ 는 다음과 같이 주어진다.

$$F_m[k] = \begin{cases} 0 & \text{for } k=0 \\ \sum_{n=0}^{W-1} f[m,n] \cos\left(\frac{\pi}{N}(n+0.5)k\right) & \text{for } k > 0 \end{cases} \quad (1)$$

위 식(1)에서 $f[m, n]$ 의 평균값은 음악의 선을 보다는 음의 크기나 음색에 더 연관되므로 평균을 차감하기 위해서 $F_m[0] = 0$ 으로 정한다. 코사인 변환의 절대값인 $|F_m[k]|$ 의 표준편차를 σ_m 라고 하면, 강인한 크로마그램 성분을 추출하기 위해서 다음과 같이 크기가 큰 값만을 남겨서 $G_m[k]$ 를 구한다.

$$G_m[k] = \begin{cases} F_m[k] & \text{if } |F_m[k]| > \gamma\sigma_m \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

식 (2)에서 γ 는 강인한 성분 추출을 위한 조정계수이다. 시간축 코사인 변환에서 크기가 큰 성분만을 남겨서 음악의 시간축 방향 음의 강약인 박자를 강조하는 것이다. 얻어진 $G_m[k]$ 를 이산 코사인 역변환하여 강인한 크로마그램 $g[m, n]$ 을 구한다. 모든 크로마그램 밴드($c=0,1,\dots,L-1$)에서 위와 같이 이산 코사인 변환, 강인한 성분 추출, 역변환을 하여 얻은 크로마그램을 $g[c, n]$ 이라고 하자. 얻어진 $g[c, n]$ 은 원래의 크로마그램인 $f[c, n]$ 처럼 매 시간 프레임 m 에 대해서 L^2 norm이 1이 되도록 정규화 한다. 얻어진 $g[c, n]$ 으로부터 Fig. 2와 같이 이차원 푸리에 변환의 크기값인 2D FTM을 구하고 중간값을 취하여 전곡 단위로 축약한다. Fig. 3은 크로마그램 블록 $f[c, n]$ 과 $g[c, n]$ 을 원본 음원과 커버곡으로 부터 각각 추출하여 도시한 예시이다. 식 (2)의 조정계수 γ 는 1의 값을 사용하였다. 시간축 코사인 변환 후 식 (2)를 통해서 크기가 작은 성분을 제거함으로, $g[c, n]$ 은 $f[c, n]$ 에 비해서 신호적으로 평탄화되지만, 주요한 화음 성분들은 잘 보존되는 것을 관찰할 수 있다. Fig. 4는 Fig. 3의 예시에서 11번째 크로마그램 밴드를 도시한 것이다. 강인한 크로마그램은 화음 성분은 보존하면서, 악기와 가수 등의 음색 차이로 인한 화음 성분의 시간적 떨림은 제거되는 효과를 보인다. 따라서 제안된 강인한 성분 추출을 통해서 원곡과 커버곡의 크로마그램 간에 신호적으로 화음 성분의 차이를 줄여주는 효과를 가지게 된다. 제안된 강인한 성분 추출 방법이 음악의 고유한 멜로디 성분은 보존하면서 음색 및 잡음 등 커버곡 검색에 도움이 되지 않는 부분의

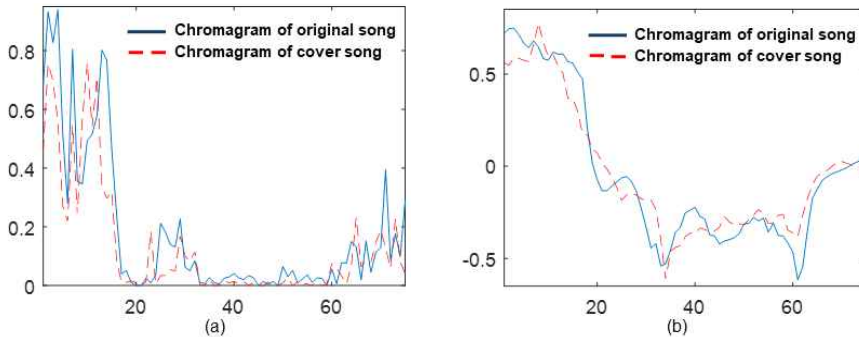


Fig. 4. (a) The 11th subband of the chromagram in Fig. 3 (a) and (b). (b) The 11th subband of the salient chromagram in Fig. 3 (c) and (d).

영향은 줄여줄 수 있다.

3. 실험 결과

기존의 2D FTM 방법[10]의 커버곡 검색 성능과 제안한 강인한 크로마그램 추출 과정을 전처리로 추가하여 성능이 개선된 2D SCFTM 방법의 성능을 비교하였다. 강인한 성분 추출을 위한 조정계수인 γ 값을 가변시켜가면서 커버곡 검색 성능을 확인하였다. 검색 성능 비교를 위해서 음원 및 성능이 공개되어 있는 covers80 데이터셋과 자체적으로 수집한 kpop100 데이터셋을 사용하였다. 미국 콜롬비아 대학에서 커버곡 실험을 위해서 수집된 covers80 데이터셋은 원본곡과 커버곡 쌍 80개로 이루어진 것으로 모두 160곡으로 구성되어있다[13]. 자체적으로 수집한 kpop100 데이터셋은 다양한 장르의 kpop으로 이루어져 있으며 원본곡과 커버곡 쌍 100개로 이루어진 것으로 모두 200곡으로 구성되어있다.

실험대상 음악들로부터 기존 2D FTM 방법[10]에서 사용된 비트 동기 크로마그램 추출 코드[14]를 사용하여 크로마 수열을 얻었다. 비트 동기 크로마그램은 비트에 동기를 맞추어 음악 템포 변화에 대해서 불변성을 얻는 것이다. 본 논문에서는 한 옥타브를 12개 구간으로 나누어 음악 비트 별로 12차 크로마 벡터를 얻었다. 이렇게 얻어진 크로마 벡터 수열을 시간축으로 S 개씩 건너뛰면서 N 개씩 모아서 블록을 구성하고 2차원 푸리에 변환의 크기값을 취한다. 본 논문에서는 2D FTM과 2D SCFTM 추출 시에 $N=75$ 와 $S=10$ 을 사용하였다. 전곡에서 얻어진 블록

들의 푸리에 변환 크기를 구하고 중간값을 구하여 전곡 특징으로 사용한다. 커버곡 음원에서 얻은 전곡 특징으로 원본 음원의 전곡특징들과 유클리디안 거리를 이용하여 비교를 수행하고, 거리가 가까운 순으로 나열하여 검색 결과를 얻는다. 커버곡 검색 결과의 MAP(mean average precision)를 성능평가지표로 사용하였다.

Figs. 5와 6은 식 (2)의 조정계수 γ 값을 0 에서 1.6 사이에서 가변하면서 얻은 커버곡 검색 정확도를 도시한 것이다. 두 실험 데이터셋 모두에서 조정계수가 너무 작거나 커지면 검색 성능이 떨어짐을 실험적으로 확인할 수 있었다. covers80 데이터셋은 γ 값이 0.8 에서 0.9사이에서 가장 좋은 성능을 보였으며, kpop 100 데이터셋은 γ 값이 0.3에서 1사이의 넓은 범위에서 좋은 성능을 보였다. 조정계수가 작으면 대부분의 크로마그램 성분들이 그대로 남게 되어 커버곡 검색에 도움이 되지 않는 음색 및 잡음 등의 영향이 제거되지 않는다. 반대로 조정계수가 커지면 음악의 고유한 정보까지 같이 제거가 되어서 크로마그램이 다른 음악들에 대해서 가지는 식별성이 떨어지게 된다. 따라서 커버곡 검색에 도움이 될 수 있도록 적당한 값의 조정계수를 선택하는 것이 중요하다. 데이터셋 별로 조금 차이가 있을 수 있지만, 실험을 통해서 조정계수 γ 값을 1보다 조금 작은 값을 사용하면 되는 것을 알 수 있었다. 조정계수 $\gamma=0$ 인 경우가 기존 2D FTM 방법과 같으며, covers80의 경우 검색 정확도가 40%였고 kpop100의 경우 30%였다. 제안된 강인한 크로마그램을 이용한 2D SCFTM의 최고 검색

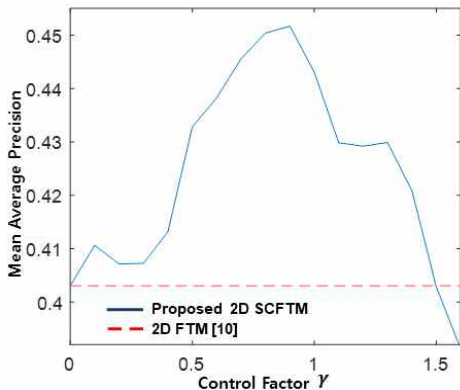


Fig. 5. Search accuracy versus control factor γ for covers80 dataset.

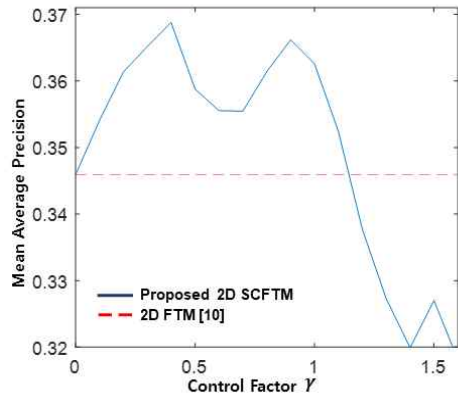


Fig. 6. Search accuracy versus control factor γ for kpop100 dataset.

정확도는 두 데이터셋에 대해서 각각 46.25%와 39%였다. 즉 제안된 2D FTMP 방법이 기존 2D FTM의 성능에 대비하여 covers80의 경우 6.25%, kpop100의 경우 9% 검색 성능을 개선함을 알 수 있다.

제안한 방법과 기존의 커버곡 검색 방법들의 성능을 비교하기 위해서, 공개된 데이터셋인 covers80에 대한 검색 성능을 Table 1에 정리하였다. 성능평가지표로써 각 커버곡을 80곡의 원곡 데이터셋과 비교하여 거리가 가까운 10개의 검색 결과가 입력 커버곡의 원곡이 맞을 경우의 확률인 P@10과 MAP를 구하였다. 두 성능평가지표 모두 큰값이 좋은 성능을 의미한다. 1장에서 살펴본 바와 같이 커버곡 검색 방법에는 전곡기반 특징 축약 방법과 크로마그램 수열 직접 비교 방법의 두가지로 나누어지며, 기존의 커버곡 검색 방법들을 그에 따라 분류하여 Table 1에 정리하였다. 각 논문에서 covers80 데이터셋에서 얻은 최고의 성능값을 Table 1에 표기하였다. 전곡기반 특징 축약 방법들 중에서 제안된 2D SCFTM은 2D FTM의 성능을 개선하나 합성곱 신경망(convolutional neural network, CNN)을 이용한 방법에 비해서는 낮은 성능을 보였다. 합성곱 신경망을 이용한 방법은 신경망 학습에 데이터와 시간이 많이 소요되며, 학습에 사용된 데이터 및 학습 방법에 따라 성능의 차이가 발생할 수 있다. 또한 제안된 방법에 비해서 최종 특징 차수도 큰 차이가 있으므로 제안된 방법과 직접적으로 성능을 비교하기 어렵다. 크로마그램 수열 직접 비교 방법들은 전곡기반 특징 축약 방법들에 비해서 우수한 성능을 보이지만 수열의 값들을 쌍을 지어서 거리 비교하는 과정이 필요하므로 검색에 소요되는 시간이 일반적으로 수백배 이상 소요되는 단점이 있다.

Table 1. Cover song search accuracy of the previous and the proposed methods on the covers80 dataset.

Method		Performance Measure	
		MAP	P@10
Song-Level Feature	2D SCFTM	0.452	0.059
	2D FTM [10]	0.403	0.057
	CNN [15]	0.509	0.068
Sequence Alignment	CRP [16]	0.544	0.061
	Fusion [17]	0.625	0.071

4. 결 론

본 논문에서는 2차원 푸리에 변환의 크기를 이용한 커버곡 검색 방법의 성능을 개선하기 위해서, 시간축 이산 코사인 변환에 기반한 강인한 크로마그램 특징 추출을 전처리로 이용하는 방법을 제안하였다. 이산 코사인 변환 상에서 크기가 큰 성분만을 남기는 방법으로 강인한 특징을 추출하였다. 제안된 전처리 과정은 커버곡 생성 과정에서 변화하지 않는 화음 성분은 보존하면서, 가수 및 악기 변화로 쉽게 바뀌는 음색 차이로 인한 변이는 줄여주는 효과가 있다. 실험을 통해서 제안한 전처리 과정을 추가하여 2차원 푸리에 변환 기반 커버곡 검색 방법의 성능을 개선할 수 있음을 보였다.

REFERENCE

[1] C. Kofler, M. Larson, and A. Hanjalic, "User Intent in Multimedia Search: A Survey of the State of the Art and Future Challenges," *ACM Computing Surveys*, Vol. 49, No. 2, pp. 1-37, 2016.

[2] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based Music Information Retrieval: Current Directions and Future Challenges," *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668-696, 2008.

[3] J. Kim and S. Cho, "Keyword Selection for Visual Search Based on Wikipedia," *Journal of Korea Multimedia Society*, Vol. 21, No. 8, pp. 960-968, 2018.

[4] P. Cano, E. Battle, T. Kalker, and J. Haitsma, "A Review of Audio Fingerprinting," *Journal of Very Large Scale Integration Signal Processing Systems for Signal, Image and Video Technology*, Vol. 41, No. 3, pp. 271-284, 2005.

[5] M. Muller and S. Ewert, "Towards Timbre-Invariant Audio Features for Harmony-based Music," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 3, pp. 649-662, 2010.

[6] M. Muller and S. Ewert, "Chroma Toolbox: Matlab Implementations for Extracting Variants

- of Chroma-based Audio Features,” *Proceeding of International Society for Music Information Retrieval Conference*, pp. 215-220, 2011.
- [7] J. Serra, E. Gomez, P. Herrera, and X. Serra, “Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 6, pp. 1138-1151, 2008.
- [8] D. Silva, C. Yeh, G. Batista, and E. Keogh, “SiMPle: Assessing Music Similarity Using Subsequences Joins,” *Proceeding of International Society for Music Information Retrieval Conference*, pp. 23-29, 2016.
- [9] T. Bertin-Mahieux and D. Ellis, “Large-Scale Cover Song Recognition Using Hashed Chroma Landmarks,” *Proceeding of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 117-120, 2011.
- [10] T. Bertin-Mahieux and D. Ellis, “Large-Scale Cover Song Recognition Using the 2D Fourier Transform Magnitude,” *Proceeding of International Society for Music Information Retrieval Conference*, pp. 241-246, 2012.
- [11] T. Kinnunen, C. Koh, L. Wang, H. Li, and E.S. Chng, “Temporal Discrete Cosine Transform: Towards Longer Term Temporal Features for Speaker Verification,” *Proceeding of International Symposium on Chinese Spoken Language Processing*, pp. 547-558, 2006.
- [12] N. Gupta, M. Swamy, and E.I. Plotkin, “Wavelet Domain-based Video Noise Reduction Using Temporal Discrete Cosine Transform and Hierarchically Adapted Thresholding,” *Institution of Engineering Technology Image Processing*, Vol. 1, No. 1, pp. 2-12, 2007.
- [13] The Covers80 Cover Song Data Set, <https://labrosa.ee.columbia.edu/projects/coverongs/cover80/> (Accessed Feb., 24, 2019).
- [14] D. Ellis and G. Poliner, “Identifying Cover Songs with Chroma Features and Dynamic Programming Beat Tracking,” *Proceeding of International Conference Acoustic, Speech and Signal Processing*, pp. 1429-1432, 2007.
- [15] X. Xu, X. Chen, and D. Yang, “Key-invariant Convolutional Neural Network Toward Efficient Cover Song Identification,” *Proceeding of IEEE International Conference on Multimedia and Expo*, pp. 1-6, 2018.
- [16] J. Serra, X. Serra, and R.G. Andrzejak, “Cross Recurrence Quantification for Cover Song Identification,” *New Journal of Physics*, Vol. 11, No. 9, pp. 1-20, 2009.
- [17] N. Chen, W. Li, and H. Xiao, “Fusing Similarity Functions for Cover Song Identification,” *Multimedia Tools and Applications*, Vol. 77, No. 2, pp. 2629-2652, 2018.



서진수

1998년 한국과학기술원 전기 및 전자공학과 공학사
 2000년 한국과학기술원 전기 및 전자공학과 공학석사
 2005년 한국과학기술원 전기 및 전자공학과 공학박사

2006년~2008년 한국전자통신연구원 선임연구원
 2008년~현재 강릉원주대학교 전자공학과 교수