

이차원 푸리에 변환의 크기와 위상을 이용한 커버곡 검색

Cover song search based on magnitude and phase of the 2D Fourier transform

서진수[†]

(Jin Soo Seo^{1†})

¹강릉원주대학교 전자공학과

(Received September 10, 2018; revised November 7, 2018; accepted November 21, 2018)

초 록: 라이브 음악 또는 리메이크를 통해서 재발매된 음악을 원곡의 커버곡이라 부른다. 본 논문은 고속 커버곡 검색을 위한 특징 축약을 위해 2차원 푸리에 변환을 이용하는 방법을 연구하였다. 이차원 푸리에 변환은 조변화에 대해서 불변성을 가지고 있으므로, 커버곡 검색을 위한 특징 축약 방법으로 적합하다. 기존 푸리에 변환 방법에서는 크기값만을 활용하였으나, 본 논문에서는 인접한 크로마 블록은 같은 조변화를 가진다는 가정하에 위상 정보를 추가로 활용하는 방법을 제안하였다. 두 가지 커버곡 실험 데이터셋에서 성능 비교를 수행하였으며, 제안된 방법이 기존 방법에 비해서 우수한 커버곡 검색 정확도를 보임을 확인하였다.

핵심용어: 음악 검색, 음악 유사도, 크로마 특징, 커버곡 판별, 2차원 푸리에 변환

ABSTRACT: The cover song refers to live recordings or reproduced albums. This paper studies two-dimensional Fourier transform as a feature-dimension reduction method to search cover song fast. The two-dimensional Fourier transform is conducive in feature-dimension reduction for cover song search due to musical-key invariance. This paper extends the previous work, which only utilize the magnitude of the Fourier transform, by introducing an invariant from phase based on the assumption that adjacent frames have the same musical-key change. We compare the cover song retrieval accuracy of the Fourier-transform based methods over two datasets. The experimental results show that the addition of the invariant from phase improves the cover song retrieval accuracy over the previous magnitude-only method.

Keywords: Music retrieval, Music similarity, Chroma feature, Cover song identification, 2D Fourier transform

PACS numbers: 43.75.Zz, 43.60.Uv

1. 서 론

컴퓨터, 스마트폰, AI 스피커 등을 비롯한 정보처리 기기, 저장매체, 네트워크 기술의 급속한 발달에 따라서 음악 데이터를 빠르고 신뢰성 있게 찾아서 제공해 줄 수 있는 음악 검색 기술의 중요성이 증대되고 있다.^[1,2] 음악 검색 서비스는 사용하는 유사도 기준에 따라서 다양하며, 핑거프린팅^[3]과 같이 입력

음악과 정확히 일치하는 아카이브상의 음악을 찾는 경우도 있고 유사음악 검색^[4,5]과 같이 특정한 성질을 공유하는 다수의 결과를 출력하는 경우도 있다. 본 논문은 핑거프린팅과 유사음악 검색 사이에 라이브 버전 및 리메이크 등을 통해서 재녹음된 음악인 커버곡을 찾는 검색 방법에 관한 것이다^[6-8]. 커버곡 검색은 오디오 아카이브 정리, 동영상공유매체를 통한 저작권 침해 방지 등에 활용될 수 있다.

커버곡은 다양한 경로를 통해서 만들어진다. 먼저 라이브 음악 녹음의 경우 악기, 가수 및 연주자의 차이로 인한 음색 변화가 두드러지게 나타나게 되고,

[†]Corresponding author: Jin Soo Seo (jsseo@gwnu.ac.kr)
Department of Electronic Engineering, Gangneung-Wonju National University, 7 Jukhun-gil, Gangneung, Gangwon-Do 25457, Republic of Korea.

(Tel: 82-33-640-2428, Fax: 82-33-656-0740)

리메이크 음악의 경우 음색변화와 함께 연주 속도 및 스타일 차이로 인한 템포 및 리듬, 음악 키의 변조, 도입부와 중심부의 위치 변화 등의 구조적 변경, 가사 변경 등이 발생한다.^[6] 이러한 다양한 변형에도 불구하고 사람들은 음악을 듣고 커버곡 유무를 어느 정도는 판별할 수 있다. 이는 커버곡 생성과정의 다양한 변형에도 불구하고 음들의 시간적 연결을 의미하는 선율(멜로디) 정보는 상대적으로 잘 보존되기 때문이다.^[9] 기존의 핑거프린팅 및 유사음악 검색에서는 주로 음악의 주파수 특성을 직접 활용하였다면, 커버곡 검색에서는 절대적인 주파수 특성이 아니라 선율 정보를 잘 표현할 수 있는 크로마 특징이 널리 사용되고 있다. 크로마는 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. 크로마 특징은 주파수 분석에 이은 옥타브 단위로 합산하는 과정에서 가수의 목소리와 악기에 연관된 음색보다는 음악 자체의 화성적 구조에 보다 밀접하게 연관되게 된다.^[9] 다른 음악 주파수 분석 방법들과 마찬가지로 크로마 특징도 음악을 프레임 단위의 짧은 구간(수십 ms ~ 수백 ms)으로 나누고, 각 프레임에서 M 차 크로마 벡터(일반적으로 $M = 12$)를 추출하게 되므로 음악 한곡으로부터 크로마 벡터 수열이 얻어지게 된다.^[10]

커버곡 검색은 크게 두 가지 방법으로 이루어진다. 두 크로마벡터 수열을 직접 비교해서 유사도를 구하는 수열 비교 방법과 크로마 벡터 수열을 추가적으로 축약하는 전곡기반 특징을 비교하여 검색하

는 방법이 있다. 수열 직접 비교 방법^[6,11]은 DNA 수열 비교에 사용된 Smith-Waterman 거리 등을 성공적으로 커버곡 검색에 적용하였으며, 검색에 시간과 계산량이 많이 소요되는 단점이 있지만 검색 성공률이 높은 장점이 있다. 전곡기반 특징 비교 방법^[12]은 검색에 필요한 특징 저장 공간을 줄일 수 있고, Fig. 1에 주어진 바와 같이 일반적으로 유클리디안 또는 해밍거리를 통한 DB 검색도 가능한 장점이 있지만 수열 직접 비교에 비해서 검색 성능이 떨어지는 단점이 있다.

본 논문은 전곡기반 특징 비교 방법 중에서 2D FTM(Fourier Transform Magnitude)을 이용한 방법^[12]을 확장한 크로마 수열 축약방법으로 2D FTMP(Fourier Transform Magnitude and Phase)를 제안한다. 커버곡 생성과정에서 음악 키의 변화로 인해서 크로마 벡터가 크로마 축으로 이동하는 경우가 자주 발생하므로, 푸리에 변환의 성질 중 푸리에 변환의 크기값의 이동 불변성을 활용한 것이 2D FTM 방법이다. 본 논문에서는 기존의 2D FTM에서 사용되지 않고 버려지는 위상정보로부터 이동 불변성을 얻는 방법을 제안하여 커버곡 검색 성능을 높이려고 한다. 푸리에 변환 위상의 경우 직접적으로 이동 불변성이 있지만, 본 논문에서는 인접 프레임 간에는 조변화의 크기가 동일하다고 가정하고 위상의 차이를 활용하여 이동 불변성을 얻는 방법을 제안하였다. 푸리에 변환 크기에 비해서 위상 정보는 상대적으로 정보량이 작지만, 본 논문처럼 푸리에 변환의 크기값과 함께 위상 정보를 같이 사용하여 성능을 개선한 결과로 음악 onset 검색^[13], 오디오 핑거프린팅^[14] 등이 있었다.

본 논문은 복소 푸리에 변환에 기반한 커버곡 검색에 관한 연구이다. II장에서 기존 2D FTM 방법을 살펴보고, 2D FTMP 방법에서 위상정보를 키변화에 불변하게 추가하는 방법을 제안한다. III장에서 제안된 방법의 성능을 실험하고 결과를 비교 분석한다.

II. 복소 푸리에 변환 기반 크로마그램 축약을 이용한 커버곡 검색

본 논문에서는 Fig. 1에 주어진 바와 같이 검색 대상 음악 파일들로부터 음악 신호를 처리하여 특징을

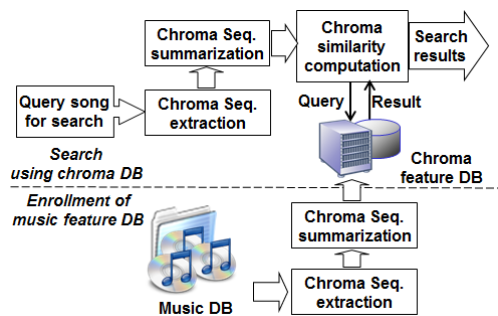


Fig. 1. Overview of the cover song search system based on song-level chromagram summarization.^[8]

추출하고 축약하여 DB에 저장하고, 입력 음악에 대해서 DB 상의 특징과의 거리 비교를 통해서 커버곡을 검색하는 방법에 관한 연구이다. 본 논문에서는 커버곡 검색을 위한 특징으로 크로마 수열을 추출하고 2차원 푸리에 변환을 통해서 조변화에 불변한 특징을 유도하는 방법을 제안한다.

2.1 2D FTM 기반 커버곡 검색 방법

2차원 푸리에 변환을 이용한 크로마 전곡 특징 추출 방법^[12]은 Fig. 2와 같다. 먼저 음악 신호로부터 크로마 벡터 수열을 추출한다. 크로마 추출 방법으로는 비트 동기 크로마^[15]를 사용한다. 비트동기 크로마는 크로마 추출과 비트 추출을 동시에 수행하고 음악 onset 사이의 크로마 값들의 평균을 취하여 출력하는 것이다. 따라서 음악의 템포 변화에 불변하게 된다. 얻어진 크로마 벡터들을 블록 단위로 묶어서 2차원 푸리에 변환을 취한다. 얻어진 푸리에 변환으로부터 음악의 조변화에 불변한 크기값을 구한다. Fig. 2에 도시한 바와 같이 각 블록의 푸리에 변환 크기값들의 중간값을 취하여 전곡 단위로 축약하게 된다. 전곡 특징에 추가적으로 PCA(Principal Component Analysis)를 적용하여 특징 차수를 줄인다.

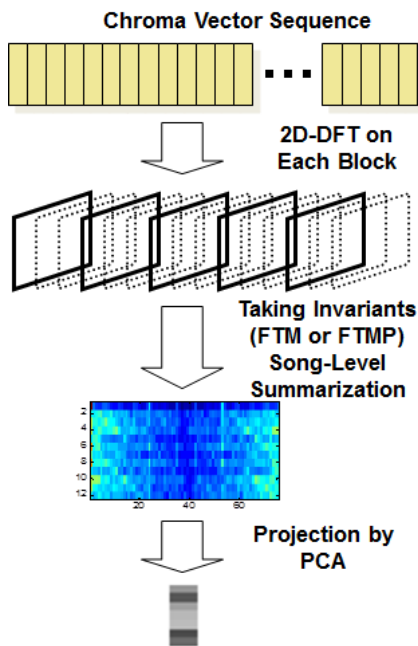


Fig. 2. Chromagram summarization using 2D Fourier transform.

2.2 2D FTMP 기반 커버곡 검색 방법

기존 2D FTM 방법은 푸리에 변환의 크기값만을 취하고 위상정보를 버리는 단점이 있었다. 본 논문에서는 버려지는 위상정보를 추가로 활용할 수 있는 방법을 제안한다. 음악의 조변화에 대한 푸리에 변환의 성질을 살펴보면 다음과 같다. 길이가 W 인 크로마 수열의 블록을 $f(c, t)$ 라고 하자. 12차 크로마를 사용할 경우 $f(c, t)$ 는 $12 \times W$ 의 크기를 가지는 2차원 신호가 된다. 음악의 조변화로 인해서 c_0 만큼 크로마 축 방향으로 신호가 이동된 커버곡의 크로마 수열 블록을 $g(c, t)$ 라고 하자.

$$g(c, t) = f(c - c_0, t). \quad (1)$$

Eq. (1) 양변을 2차원 푸리에 변환하면 f 와 g 의 푸리에 변환인 $F(\xi, \eta)$ 와 $G(\xi, \eta)$ 간에 다음과 같은 관계를 얻을 수 있다.

$$G(\xi, \eta) = e^{-j2\pi\xi c_0} F(\xi, \eta). \quad (2)$$

Eq. (2)로부터 $F(\xi, \eta)$ 와 $G(\xi, \eta)$ 의 크기값은 동일하며, c_0 만큼 이동함에 따라서 위상차이가 발생함을 알 수 있다. 이러한 2차원 푸리에 변환의 신호 이동에 대한 불변성은 영상 인식에 널리 사용되어왔다.^[16]

Eq. (2)에 주어진 신호 이동에 대한 푸리에 변환의 크기와 위상 불변성을 크로마 수열에 적용하도록 하겠다. 음악 신호로부터 얻은 크로마 벡터들을 얻어진 시간순으로 크로마 벡터 수열 $\mathbf{X} = (X_1, X_2, \dots, X_N)$ 로 표기하면, S 프레임 만큼 이동하면서 크로마 벡터 수열을 W 개씩 묶어서 크로마 벡터 블록을 만든다. 크로마 벡터 블록은 \mathbf{X} 의 부분 수열이 되며, m 번째 블록인 $(X_{(m-1)S+1}, X_{(m-1)S+2}, \dots, X_{(m-1)S+W})$ 신호에 $12 \times W$ 크기의 2차원 푸리에 변환을 취한다. 이렇게 \mathbf{X} 로부터 구해지는 L 개의 블록들에 각각 2차원 푸리에 변환을 취해서 $\mathbf{F} = (F_1, F_2, \dots, F_L)$ 를 얻었다고 하자. Eq. (2)에서 알 수 있듯이 F_i 들의 크기값은 음악 조변화에 무관하므로 그대로 사용한다. 위상값의 변화를 반영해 주기 위해서 인접한 블록은 같은 조변화를 겪으므로 크로마 축 이동값 c_0 가 같다고 가정한다. 그러면 인접한 블록의 푸리에 변환 위상간에는 Eq. (2)와 같이

같은 값만큼 위상 변화가 발생한다. 다음 Eq. (3)과 같이 F_{i+1} 위상값에서 F_i 의 위상값을 빼주면 크로마 축 방향 이동에 불변하는 신호 H_i 를 얻을 수 있다.

$$H_i = |F_i| \exp(j(\theta_{F_{i+1}} - \theta_{F_i})). \quad (3)$$

따라서 조변화에 따른 불변성을 유지하면서도 푸리에 변환의 크기 뿐만 아니라 위상 정보까지 커버곡 검색에 활용할 수 있다. Fig. 3은 커버곡 쌍에 대해서 얻어진 불변신호 H_i 를 도시한 예시이다. Fig. 3의 예시는 Fig. 3(a)와 (b)의 크로마그램에서 알수 있듯이 크로마축으로 2 만큼 이동(조변화)되었음을 알 수 있다. Fig. 3(c)와 (d)는 각각 H_i 의 실수부와 허수부

이다. 원본 음악으로부터 얻은 H_i 와 커버곡의 H_i 가 유사함을 알 수 있다. 불변신호 임에도 차이가 나는 이유는 커버곡 생성과정에서 가수 및 악기 변경으로 인한 음색차이 때문이다. Fig. 3(e)와 (f)는 커버곡 관계가 아닌 두 곡으로부터 얻은 H_i 의 실수부와 허수부이다. 상대적으로 차이가 더 큰 것을 확인할 수 있다. 마지막으로 Eq. (3)의 H_i 는 복소수이므로 실제로는 실수부와 허수부를 각각 구하고 Fig. 2에 주어진 바와 같이 시간축 방향으로 평균을 구하여 전곡 특징을 얻는다. 제안된 2D FTMP에서도 기존의 2D FTM과 마찬가지로 전곡 특징에 PCA를 적용하여 특징 차수를 줄인다.

III. 실험 결과

본 장에서는 II장에서 제안한 2D FTMP 와 기존의 2D FTM 방법의 커버곡 검색 성능을 비교한다. 특징 추출 파라미터인 블록 길이 W 와 PCA 차수를 가변시켜가면서 커버곡 인식 성능을 확인하였다. 커버곡 성능 비교를 위해서 음원 및 성능이 공개되어 있는 covers80 데이터셋과 자체적으로 수집한 kpop100 데이터셋을 사용하였다. 미국 콜롬비아 대학에서 커버곡 실험을 위해서 수집된 covers80 데이터셋은 원본 곡과 커버곡 쌍 80개로 이루어진 것으로 모두 160곡으로 구성되어있다.^[17,18] 자체적으로 수집한 kpop100 데이터셋은 다양한 장르의 kpop으로 이루어져 있으며 원본곡과 커버곡 쌍 100개로 이루어진 것으로 모두 200곡으로 구성되어있다.

실험대상 음악들로부터 기존 비교 대상 방법^[12]에서 사용된 비트 동기 크로마 추출 코드^[15]를 사용하여 크로마 수열을 얻었다. 비트 동기 크로마는 비트에 동기를 맞추어 음악 템포 변화에 불변하게 된다. 본 논문에서는 한 옥타브를 12개 구간으로 나누어 음악 비트 별로 12차 크로마 벡터를 얻었다. 이렇게 얻어진 크로마 벡터 수열을 시간축으로 S 개씩 건너 뛰면서 W 개씩 모아서 블록을 구성하고 2차원 푸리에 변환을 취한다. 기존 2D FTM은 크기만을 사용하고 제안된 2D FTMP는 크기와 위상정보를 모두 사용하여 전곡 특징을 얻었다. 원본 음원과 커버곡 음원으로부터 얻은 전곡 특징 간의 유클리디안 거리를

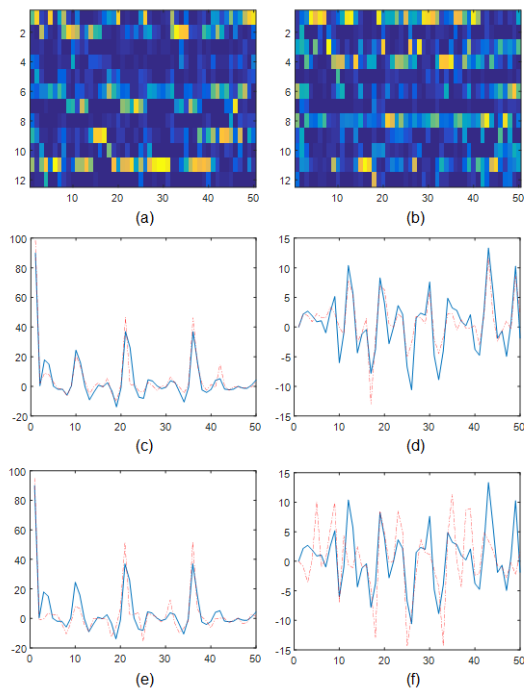


Fig. 3. (a) Chromagram of the excerpt of the original song "Between the bars". (b) Chromagram of the excerpt of the cover song "Between the bars". (c) Real part of H_i from (a) and (b) is given by solid and dashed line respectively. (d) Imaginary part of H_i from (a) and (b) is given by solid and dashed line respectively. (e) Real part of H_i from (a) and another song ("My heart will go on") is given by solid and dashed line respectively. (f) Imaginary part of H_i from (a) and another song ("My heart will go on") is given by solid and dashed line respectively. From (c) to (f), first 50 coefficients of zigzag scan of H_i are displayed (i.e. low-frequency components).

구하여 원본 음원과 다른 DB내 음원들의 전곡 특징 간의 유클리디안 거리들보다 작을 경우 커버곡 검색이 성공한 것으로 한다. 데이터셋에 있는 각 원본 음원에 대해서 데이터셋 내의 음원들과 거리 비교를 수행하고 가장 거리값이 작은 것이 커버곡일 확률을 검색 정확도로 사용하였다.

블록 길이 W 를 가변하면서 얻은 커버곡 검색 성능을 Figs. 4와 5에 도시하였다. 스텝사이즈 S 는 10으로 고정하였으며, PCA를 통해서 축약하지 않고 전곡 특징을 그대로 사용해서 얻은 결과이다. 먼저 블록 길이값이 커짐에 따라서 검색 성능이 좋아지는 경향성을 보였으나 블록 길이가 50 이상이 될 경우 성능에 큰 변화가 없었다. 제안된 2D FTMP 방법이 기존 2DFTM의 성능을 개선함을 알 수 있으며, 특히 블록의 크기가 작을 경우 성능 개선의 정도가 더 크다. 제

안된 2D FTMP 방법이 기존 2D FTM의 성능에 대비하여 covers80의 경우 2%, kpop100의 경우 8% 정도 검색 성능을 개선함을 알 수 있다.

PCA를 전곡 특징에 적용하여 특징 차수를 줄인 후 커버곡 검색을 수행하고 얻은 커버곡 검색 성능을 Figs. 6과 7에 도시하였다. PCA 축약을 위해서 실험에 사용된 데이터셋과 겹치지 않도록 1,500곡의 음악으로부터 2DFTM과 2DFTMP의 전곡 특징을 구하고 고유값 해석을 통해서 투영 행렬을 구하였다. 얻어진 투영 행렬을 이용하여 PCA 차수를 10에서 80까지 변화시켜가면서 커버곡 검색 성능을 구하였다. Figs. 4와 5의 PCA로 축약하지 않고 실험한 결과와 비교해보면 PCA를 통해 특징을 축약하더라도 특징 차수를 50 이상 유지하면 검색 성능이 비슷함을 알 수 있다. 기존 2D FTM에 비해서 2D FTMP는 위상 정보

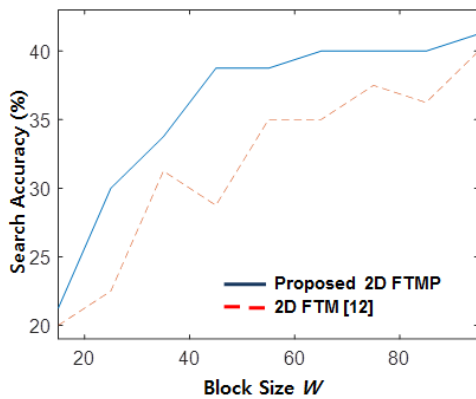


Fig. 4. Search accuracy (%) versus block size W for covers80 dataset.

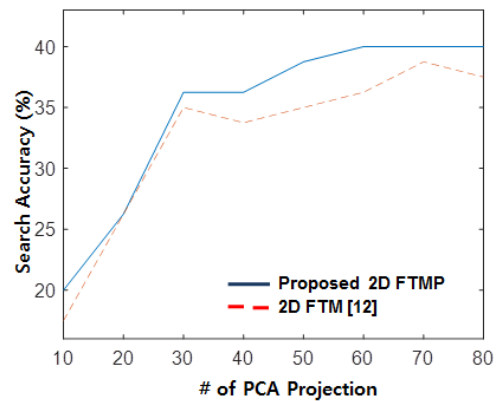


Fig. 6. Search accuracy (%) versus PCA dimension for covers80 dataset with $W = 75$.

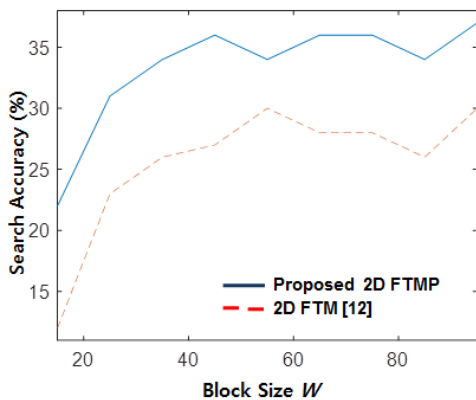


Fig. 5. Search accuracy (%) versus block size W for kpop100 dataset.

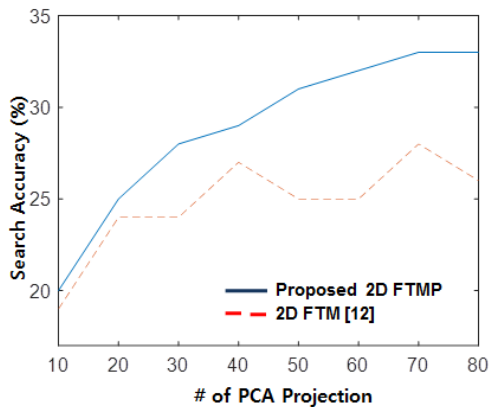


Fig. 7. Search accuracy (%) versus PCA dimension for kpop100 dataset with $W = 75$.

추가로 인해서 복소수 형태로 주어지므로, 실수부와 허수부를 따로 고려해서 특징 차수가 2배로 늘어나는 단점이 있다. 하지만 Figs. 6과 7의 결과는 PCA를 취한 이후의 검색 성능을 보기 위해서 2D FTMP와 2D FTM의 PCA 이후 최종 특징 차수를 동일하게 했을 때 성능 비교이다. 즉 PCA 이후 최종 전곡 특징 차수를 2D FTMP와 2D FTM 같은 값으로 할 경우에도 본 논문에서 제안한 2D FTMP가 기존 2D FTM의 검색 성능을 개선함을 알 수 있다. 즉, 위상 정보 추가로 인해서 2D FTMP의 차수는 2D FTM의 두배가 되지만, PCA로 축약하여 같은 차수로 맞추고 성능을 비교 하더라도 제안된 2D FTMP가 실험대상 두 데이터셋 모두에서 비슷하거나 더 우수한 검색 정확도를 보였다.

공개된 covers80 데이터셋에 대해서 수열을 직접 비교하는 방법들^[6,15]의 검색 정확도와 속도를 2D FTM과 2D FTMP 방법과 비교하였다. 성능 비교는 인텔 CPU i7-4790(3.6 GHz), RAM 16 GB 컴퓨터 상에서 MATLAB 소프트웨어를 사용하였다. 특징 추출 및 변환 시간을 제외한 covers80 데이터셋에서 순수하게 검색에 소요된 총시간(즉, 크로마 수열간의 거리 비교 시간)과 검색 결과의 정확도를 구하였다. 생체 신호 수열비교 방법인 Smith-Waterman 거리를 이용한 방법^[6]은 80곡에 대한 검색 실험(80곡 입력, 80곡 커버곡 DB 이므로 총 6400 음악쌍 비교)을 수행하는데 637 s 소요되었으며 검색정확도는 57.5%였다. 크로마수열을 상호상관을 통해서 직접 비교한 방법^[15]은 80곡 DB에 대한 검색 실험에 38 s 소요되었으며 검색정확도는 43.8%였다. 수열을 전곡특징으로 축약하는 2D FTMP 방법은 PCA로 차수를 줄이지 않더라도 80곡 DB에 대해서 0.015 s의 총검색 시간이 소요되었으며 PCA로 압축할 경우 검색시간을 더 줄일 수 있다. Fig. 4에 도시한 바와 같이 최고 검색정확도는 41.3%였다. 이와 같이 수열을 직접 비교하는 방법들^[6,15]과 비교하면 2D FTM과 2D FTMP 방법의 검색 정확도가 상대적으로 우수하지는 않다. 하지만 검색에 소요되는 시간 측면에서 크게 이점이 있다. 검색 소요 시간 외에도 2D FTM과 2D FTMP 방법은 전곡 기반으로 특징을 축약하고 추가적으로 PCA를 통해서 특징 차수를 줄이므로 특징 DB의 크기도 줄어들어 저장공간을 절약하는 이점도 있다. 또한 유클리디안

거리를 통한 DB 검색이 용이한 장점이 있어서 대규모 음악 DB에서 실제로 커버곡 검색 서비스에 적용 가능한 방법이다.

IV. 결 론

본 논문에서는 2차원 푸리에 변환의 크기와 위상을 이용하여 음악 조변화에 불변한 특징을 제안하고 커버곡 검색에 적용하였다. 기존 연구에서는 푸리에 변환의 불변성을 활용하기 위해서 크기값만을 사용하였으나, 본 논문에서는 인접 프레임 간에는 조변화의 크기가 동일하다고 가정하고 위상의 차이를 활용하여 이동 불변성을 얻는 방법을 제안하였다. 위상 정보의 추가로 인해서 특징 차수가 커지는 단점이 있지만, PCA를 통해서 특징 차수를 줄일 경우 같은 차수에서도 제안된 2D FTMP 방법이 기존 2D FTM 방법 대비 검색 성능이 우수함을 실험으로 확인하였다. 커버곡 검색을 위한 푸리에 변환 기반 특징 축약 방법에서 크기와 함께 위상 정보를 추가적으로 활용하여 성능 개선을 이룰 수 있음을 확인하였으며, 음악 장르 분류 등 다른 음악 정보 처리에도 위상 정보의 활용 필요성에 대한 연구가 필요할 것으로 생각된다.

감사의 글

이 논문은 2018년도 강릉원주대학교 학술연구조성비 지원에 의하여 수행되었음.

References

1. Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. Multimedia* **13**, 303-319 (2011).
2. J. Seo, J. Kim, and J. Park, "Centroid-model based music similarity with alpha divergence" (in Korean), *J. Acoust. Soc. Kr.* **35**, 83-91 (2016).
3. J. Lee and H. Kim, "Audio fingerprinting using a robust hash function based on the MCLT peak-pair" (in Korean), *J. Acoust. Soc. Kr.* **34**, 157-162 (2015).
4. B. Logan and A. Salomon, "A music similarity function based on signal analysis," *Proc. ICME-2001*,

- 745-748 (2001).
5. C. Charbuillet, D. Tardieu, and G. Peeters, "GMM supervector for content based music similarity," Proc. DAFX-2011, 425-428 (2011).
 6. J. Serra, E. Gomez, P. Herrera, and X. Serra, "Chroma binary similarity and local alignment applied to cover song identification," IEEE Trans. Audio Speech Lang. Process. **16**, 1138-1151 (2008).
 7. P. Foster, S. Dixon, and A. Klapuri, "Identifying cover songs using information-theoretic measures of similarity," IEEE Trans. Audio Speech Lang. Process. **23**, 993-1005 (2015).
 8. J. Seo, J. Kim, and J. Park, "An investigation of chroma n-gram selection for cover song search" (in Korean), J. Acoust. Soc. Kr. **36**, 436-441 (2017).
 9. M. Muller and S. Ewert, "Towards timbre-invariant audio features for harmony-based music," IEEE Trans. Audio Speech Lang. Process. **18**, 649-662 (2010).
 10. M. Muller and S. Ewert, "Chroma toolbox: MATLAB implementations for extracting variants of chroma-based audio features," Proc. ISMIR-2011, 215-220 (2011).
 11. D. Silva, C. Yeh, G. Batista, and E. Keogh, "SIMPLE: Assessing music similarity using subsequences joins," Proc. ISMIR-2016, 23-29 (2016).
 12. T. Bertin-Mahieux and D. Ellis, "Large-scale cover song recognition using the 2D Fourier transform magnitude," Proc. ISMIR-2016, 241-246 (2012).
 13. J. Bello, C. Duxbury, M. Davies, and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," IEEE Signal Process. Letters **11**, 553-556 (2004).
 14. J. Seo, J. A. Haitsma, and T. Kalker, "Linear speed-change resilient audio fingerprinting," Proc. MPCA-2002, 45-48 (2002).
 15. D. Ellis and G. Poliner, "Identifying cover songs' with chroma features and dynamic programming beat tracking," Proc. ICASSP-2007, 1429-1432 (2007).
 16. B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," IEEE Trans. Image Process. **5**, 1266-1271 (1996).
 17. *The covers80 cover song data set, available*, <https://labrosa.ee.columbia.edu/projects/cover songs/covers80/>, 2007.
 18. D. Ellis and C. Cotton, "The 2007 LabROSA cover song detection system," in MIREX extended abstract 2007, (2007).

저자 약력

▶ 서진수 (Jin Soo Seo)



1998년 2월: KAIST 전기 및 전자공학과 공학사
 2000년 2월: KAIST 전기 및 전자공학과 공학석사
 2005년 2월: KAIST 전기 및 전자공학과 공학박사
 2006년 3월 ~ 2008년 2월: 한국전자통신연구원 선임연구원
 2008년 3월 ~ 현재: 강릉원주대학교 전자공학과 교수