한국음향학회지 제41권 제6호 pp. 662~667 (2022)
The Journal of the Acoustical Society of Korea Vol.41, No.6 (2022)
https://doi.org/10.7776/ASK.2022.41.6.662

pISSN : 1225-4428
eISSN : 2287-3775

커버곡 검색을 위한 확률적 선형 판별 분석 기반 음악 유사도

A music similarity function based on probabilistic linear discriminant analysis for cover song identification

서진수,[1†] 김정현,[2] 김혜미[2]

(Jin Soo Seo,[1†] Junghyun Kim,[2] and Hyemi Kim[2])

[1]강릉원주대학교 전자공학과, [2]한국전자통신연구원 콘텐츠연구본부

(Received September 30, 2022; accepted October 27, 2022)

초 록: 음악 유사도 계산은 음악 검색 서비스 구현에서 가장 중요한 요소 중 하나이다. 본 논문은 커버곡 검색의 성능을 제고하기 위한 음악 유사도 학습에 대해서 다룬다. 음악 유사도 함수를 유도하는 데 확률적 선형 판별 분석을 이용하여 잠재 음악 공간을 구한다. 잠재 음악 공간은 같은 커버곡 간의 거리는 줄이고 다른 곡 간의 거리는 크게 되도록 학습한다. 추출된 음악 특징이 잠재 음악 변수에서 생성되었다는 가정 하에 확률 모델을 구하고, 음악의 동질성 여부를 가설검증하여 음악 유사도 함수를 유도한다. 두 가지 커버곡 실험 데이터셋에서 성능 비교를 수행하여 제안한 음악 유사도 함수가 커버곡 검색 성능을 개선시킬 수 있음을 보였다.

핵심용어: 커버곡 검색, 음악 유사도, 확률적 선형 판별 분석, 잠재 변수

ABSTRACT: Computing music similarity is an indispensable component in developing music search service. This paper focuses on learning a music similarity function in order to boost cover song identification performance. By using the probabilistic linear discriminant analysis, we construct a latent music space where the distances between cover song pairs reduces while the distances between the non-cover song pairs increases. We derive a music similarity function by testing hypothesis, whether two songs share the same latent variable or not, using the probabilistic models with the assumption that observed music features are generated from the learned latent music space. Experimental results performed on two cover music datasets show that the proposed music similarity improves the cover song identification performance.

Keywords: Cover song identification, Music similarity, Probabilistic Linear Discriminant Analysis (PLDA), Latent variable

PACS numbers: 43.75.Zz, 43.60.Uv

I. 서 론

정보처리 기기와 유무선 네트워크 기술의 발달에 따라서 사용자가 원하는 각종 콘텐츠를 빠르고 신뢰성 있게 찾아서 제공해 줄 수 있는 검색 기술의 필요성이 커지고 있다. 음원 유통 서비스와 관련하여, 사용자의 요구에 맞추어 음악을 찾아서 제공하는 것을 가능하게 하는 음악 정보 처리 및 검색 기술이 널리 연구되고 있다.[1-3] 본 논문은 다양한 음악 검색 문제 중에서 커버곡 검색의 성능을 높이기 위한 음악 유사도 비교 방법에 관해서 다룬다. 커버곡은 콘서트 현장에서 라이브 녹음, 편집이나 리메이크 등을 통해서 재녹음된 음악을 가리킨다.[3] 커버곡 검색 기술은 웹하드 및 유튜브 등 데이터 공유 서비스에서 저작권 보호, 중복된 음원을 가진 음악 아카이브 정리 등에 활용될 수 있을 것으로 기대된다.

†**Corresponding author:** Jin Soo Seo (jsseo@gwnu.ac.kr)
Department of Electronic Engineering, Gangneung-Wonju National University, 7 Jukhun-gil, Gangneung, Gangwon-Do 25457, Repubic of Korea
(Tel: 82-33-640-2428, Fax: 82-33-656-0740)

음악 검색은 목적에 따라서 다양한 유사도 기준을 가지고 있다. 예를 들어 핑거프린팅은 입력 음악과 정확히 일치하는 아카이브상의 음악을 찾고, 유사음악 검색 서비스는 같은 장르의 비슷한 음색을 공유하는 다수의 결과를 출력한다. 음악 검색 목적에 따라서 사용하는 특징과 음악 유사도 비교 방법이 달라진다. 다양한 음악 검색 문제 중에서 본 논문은 커버곡 검색을 다루며, 특히 커버곡 검색을 위한 음악 유사도 비교 방법에 관해서 연구한다. 커버곡 검색을 위한 음악 유사도 함수를 찾기 위해서는 원곡과 커버곡 간의 공유되는 특성을 찾아야 한다. 하지만 커버곡을 만드는 과정에서 실황 녹음, 편집, 리메이크를 거치면서, 가수와 악기의 차이로 인한 음색 변화, 연주 속도 및 스타일 차이로 인한 템포, 리듬, 음악 키 변조 등 원곡과 커버곡 간에 다양한 종류의 변형이 발생한다.[3-5] 이런 다양한 변형이 존재하므로 커버곡 검색을 위한 음악 특징을 찾고 유사도 함수를 정의하는 것은 여전히 어려운 문제로 남아있다.

커버곡 검색을 위한 음악 특징의 경우 대표적으로 화음과 멜로디를 표현하는 특징인 크로마그램이나 Constant-Q Transform(CQT) 등이 사용되고 있다.[3] 일반적으로 커버곡 검색은 전곡 단위 입력에 대해서 이루어지며, 음악 신호로부터 얻어지는 특징 수열을 직접 비교하여 시간축 상에서 정합하는 수열 직접 비교 방법[4]과 특징 벡터 수열을 가공하여 검색에 용이한 고정된 길이의 전곡 특징[5]을 구하는 방법으로 나눌 수 있다. 수열 직접 비교 방법은 음성 인식과 DNA 수열 분석에서 사용해왔던 특징 비교 방법을 사용하여 우수한 검색 정확도를 보이는 장점이 있으나, 수열 비교에 많은 계산량이 요구되고 특징 벡터 수열을 모두 저장해야 하므로 저장 공간이 많이 필요하다. 반면 전곡 특징 추출 방법은 특징 축약 과정에서 유실되는 정보로 인해서 커버곡 검색 성능을 향상시키는 것에 한계를 보였으나, 최근 딥러닝을 전곡 특징 추출에 적용하여 검색 정확도가 크게 제고되었다.[6-8] 본 논문은 Fig. 1에 주어진 바와 같이 추출된 전곡 특징에 확률적 선형 판별 분석 Probabilistic Linear Discriminant Analysis(PLDA)[9-11]를 이용한 음악 유사도 비교 방법을 제안한다. PLDA는 얼굴 인식에 처음 적용되었으며,[9] 이후 화자 인식에 널리

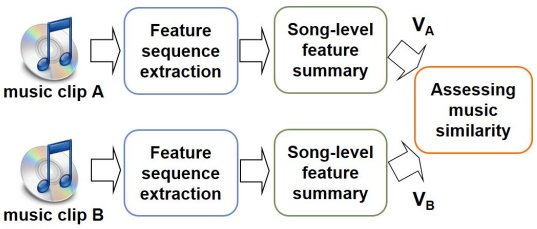

Fig. 1. (Color available online) Overview of the music similarity computation for cover song identification based on song-level feature summary.

사용되었다.[10] 특히 딥러닝 모델 기반 화자 인식에서도 유사도 비교에 적용되어 우수한 성능을 보였다.[11] 본 논문에서는 PLDA를 커버곡 검색을 위한 음악 유사도 함수를 구하기 위해 사용하였다. PLDA를 통해서 같은 커버곡 간의 특징 변이와 다른 음악들 간의 특징 변이를 각각 확률 모델로 표현한다. Fig. 1에 도시한 바와 같이 두 음악 신호로부터 추출한 음악 전곡 특징 V_A와 V_B를 미리 학습한 특징 변이 확률 모델에 적용하여, V_A와 V_B 간의 특징 변이가 같은 커버곡 간의 변이에 속하는 지 다른 음악들 간의 변이에 더 가까운 지 우도를 계산하여 커버곡 관계 여부를 판정한다. PLDA를 통해서 특징 변이 확률 모델이 커버곡 변이 패턴을 학습하게 되어서 PLDA 기반 음악 유사도는 커버곡 검색 성능을 제고할 수 있다.

II. PLDA 기반 커버곡 유사도 비교

본 논문은 Fig. 1에 주어진 음악 전곡 특징 기반 커버곡 검색에 관해서 다룬다. 커버곡 검색에 사용되는 음악 전곡 특징으로 Musically-Motivated Version Embeddings(MOVE) 모델[7]에 대해서 살펴보고, MOVE 출력을 PLDA를 통해서 확률 모델링하고 음악 유사도를 유도하여 커버곡 검색 성능을 개선하는 방법을 제안한다. 기존 커버곡 검색 방법들에서는 두 전곡 특징 간의 거리를 직접 계산하여 음악 유사도로 이용한 것에 반해서, 제안된 방법은 두 전곡 특징이 같은 곡인지 다른 곡인지 가설 검증하고 우도를 기반으로 음악 유사도를 구한다.

2.1 커버곡 검색을 위한 전곡 특징 추출

커버곡 검색은 음악 신호 전체에서 얻은 특징에

대해서 이루어지며, Fig. 1에 도시한 바와 같이 특징을 축약하여 전곡 특징을 얻고 유사도를 비교한다. 본 논문에서는 특징 추출을 위한 딥러닝 모델 코드 및 미리 학습된 파라미터 값이 제공되는 MOVE를 음악 전곡 특징 추출 방법으로 사용하였지만, 2.2절에서 제안하는 PLDA 기반 음악 유사도 학습 방법은 다른 음악 전곡 특징에도 적용할 수 있다. Fig. 2에 도시한 바와 같이 음악 신호를 93 ms 길이 프레임으로 나누고 프레임 당 12차의 Convolutional and Re-current Estimators for Music Analysis(CREMA) 특징[12]을 구하고 신경망 입력으로 사용한다. CREMA 특징은 음악의 코드 정보를 표현하도록 학습되어서 다른 크로마그램들에 비해서 우수한 커버곡 검색 성능을 보였다.[7] MOVE 모델의 첫 번째 레이어는 음악 키의 변조에 강인할 수 있도록 주파수축 이동에 불변하는 신경망 구조를 사용한다. 두 번째부터 다섯 번째 레이어 까지 4개의 컨벌루션 레이어는 시간축 방향으로만 컨벌루션을 수행한다. 이렇게 1개의 주파수축 불변 레이어와 4개의 시간축 컨벌루션 레이어를 통해서 음악 신호의 광역 시간-주파수 정보를 얻는다. 마지막으로 컨벌루션 레이어 출력 채널의 주의집중 가중치를 학습하고 곱한 후에 채널 별로 합산하여 전역 풀링(global pooling)을 수행한다. 전역 풀링 후 얻은 256차 특징을 16000개의 출력 노드를 가진 선형 레이어에 통과시켜서 16000차 특징을 얻는다. MOVE 모델은 8만3천곡 규모의 커버곡 데이터셋에서 삼중항 손실(triplet loss)[13]을 최소화하도록 학습하여 구해지며 학습된 파라미터 값이 공개되어 있다. MOVE 모델의 자세한 구조 및 상세한 학습 방법은 Reference [7]에 있다.

기존 MOVE 방법의 전곡 특징 차수는 16000으로 실제 활용하기에 너무 크기 때문에, 본 논문에서는 Batch Normalization NECK(BNNECK) 구조[14]를 이용하여 특징을 축약하였다. Fig. 2에 도시한 바와 같이 기존 MOVE 모델의 전역 풀링층과 최종 출력단 사이에 배치 정규화 레이어를 추가하고 재학습하여 256차의 BNNECK 출력 특징을 구한다. BNNECK 특징은 신경망의 전역 풀링층과 최종 출력단 사이에 위치하여 커버곡 인식에 유용한 정보가 축약된다. BNNECK 구조를 통해서 음악 한곡을 256차 특징으로 표현하고, 2.2절에서 얻어진 256차 전곡 특징에 PLDA를 적용하여 음악 유사도를 구한다.

2.2 PLDA를 이용한 음악 유사도

PLDA는 관찰된 대상들 마다 정체성을 표현하는 잠재 변수가 존재한다는 가설에 기반한다.[9] 같은 정체성을 가지더라도 다른 형태로 관찰될 수 있지만, 이들 관찰된 대상들의 정체성을 나타내는 잠재 변수는 동일하다는 가정이다. 이 PLDA 가정을 커버곡 검색에 적용하면 음악 신호는 연주 때마다 가수 또는 악기에 따라서 다르게 출력될 수 있지만 만약 동일한 음원이라면 같은 값의 잠재 변수를 가져야 한다. Fig. 3에 주어진 바와 같이 관찰된 음악 신호가 **A**, **A'**, **B**, **C**의 4가지 라고 할 때, **A**와 **A'**이 커버곡 관계에 있다면 **A**와 **A'**의 잠재 변수값은 PLDA를 통해서 $\boldsymbol{h}_A$로 같은 값을 가지게 된다. 즉, 성공적인 커버곡 검색을 위해서는 **A**와 **A'** 사이에 가사, 악기, 음악 키 변조, 연주 속도 등 다양한 차이가 존재하더라도 이상적으로

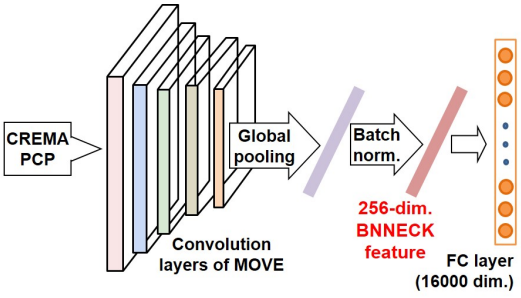

Fig. 2. (Color available online) Song-level feature summarization using musically-motivated version embeddings.[7]

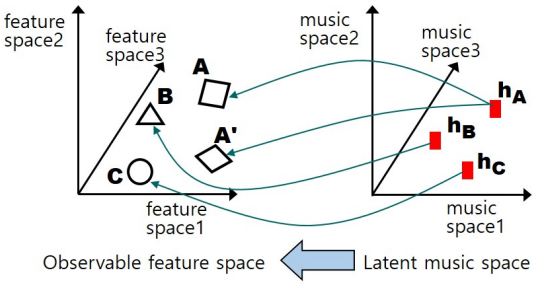

Fig. 3. (Color available online) Conceptual diagram of the latent music space which generates observable features. The h_A in the latent space generates both A and its cover A'.

잠재 변수 $\boldsymbol{h}_A$는 같은 값을 가지게 되도록 잠재 음악 공간을 학습한다.

본 논문에서는 Fig. 3의 관찰된 음악 **A**, **A'**, **B**, **C**를 2.1절에서 살펴본 256차원 MOVE 전곡 특징이라고 하고, 커버곡 학습 데이터를 통해서 커버곡 간의 잠재 변수 $\boldsymbol{h}$값의 차이가 작도록 PLDA 모델을 학습한다. 하나의 음원으로부터 R개의 커버곡이 생성되었고, 생성된 커버곡들의 MOVE 특징을 각각 V_1, V_2, ..., V_R이라고 하자. 표준 정규분포 기반 PLDA를 사용하면 각각의 MOVE 특징 V_r을 전체 평균 $\boldsymbol{m}$과 해당 음원의 정체성을 나타내는 잠재 변수인 $\boldsymbol{h}$(표준 정규분포 가정), 같은 커버곡 간의 특징 변이로는 평균 0이고 공분산 Σ가 완전위수(full rank)인 ε_r로 다음과 같이 표현할 수 있다.

$$\begin{aligned} V_r &= m + Sh + \epsilon_r \\ where\ \ h &\sim N(0, I) \\ \epsilon_r &\sim N(0, \Sigma) \end{aligned} \tag{1}$$

Eq. (1)에서 PLDA 모델 파라미터인 m, S, Σ는 커버곡 학습 데이터로부터 커버곡 변이 정보를 바탕으로 Expectation Maximization(EM) 알고리즘으로 학습한다.[9]

얻어진 PLDA 모델을 이용하여 MOVE 특징들이 커버곡 관계에 있는지 여부를 판별한다. Fig. 1에서 음악 A와 B의 MOVE 전곡 특징을 각각 V_A와 V_B 라고 하면, V_A와 V_B 가 음악 정체성을 나타내는 잠재 변수 값이 같은 지를 확인하면 A와 B가 커버곡 관계 인지 여부를 판별할 수 있으므로 다음과 같은 가설 검증을 유도하였다.

가설 L_0: 두 음악 A와 B가 같은 음원을 공유하는 커버곡 관계에 있어서, V_A와 V_B 가 같은 잠재 변수 $\boldsymbol{h}$값을 가진다.

가설 L_1: 두 음악 A와 B가 다른 음원이어서, V_A와 V_B 가 각각 다른 잠재 변수 $\boldsymbol{h}$값을 가진다.

위 가설 검증을 통해서 두 음악 A, B간 유사도 $\boldsymbol{s}$(A,B)는 다음과 같이 로그 우도비로 표현할 수 있다.

$$s(A,B) = \log \frac{p(V_A, V_B|L_0)}{p(V_A|L_1)p(V_B|L_1)}. \tag{2}$$

Eq. (1)의 정규분포 가정을 (2)에 적용하면 음악 유사도 $\boldsymbol{s}$를 닫힌 해(closed-form solution)로 구할 수 있으며 다음과 같이 주어진다.

$$\begin{aligned} s(A,B) = &\log \mathcal{N}\left(\begin{bmatrix} V_A \\ V_B \end{bmatrix}; \begin{bmatrix} m \\ m \end{bmatrix}, \begin{bmatrix} \Sigma + SS^T & SS^T \\ SS^T & \Sigma + SS^T \end{bmatrix} \right) \\ &- \log \mathcal{N}\left(V_A; m, \Sigma + SS^T \right) \\ &- \log \mathcal{N}\left(V_B; m, \Sigma + SS^T \right) \end{aligned} \tag{3}$$

PLDA 모델 학습 및 로그 우도 유도 과정은 References [9]와 [10]에 자세히 기술되어 있다.

III. 실험 결과

제안한 PLDA 기반 음악 유사도를 MOVE에 적용하여 커버곡 검색 성능을 확인하였다. 커버곡 검색 성능 확인을 위해서 MOVE에서 입력으로 사용하는 CREMA 특징이 제공되는 Da-tacos 테스트 데이터셋[15]과 음원이 공개되어 있어서 커버곡 검색 성능 비교를 위해서 널리 사용되어온 covers80 데이터셋[16]을 사용하였다. Da-tacos 테스트 데이터셋은 음원이 제공되지 않고 추출된 특징과 메타데이터만 제공되며, 전체 1만 5천곡 중에서 1만 3천곡은 13개의 커버 버전으로 이루어진 1000개 음원이며 나머지 2천곡은 검색 성능을 평가하기 위해서 사칭자(imposter)로 삽입되었으며 실험에서 검색 입력으로 사용되지 않았다. 미국 콜롬비아 대학에서 커버곡 실험을 위해서 수집된 covers80 데이터셋은 원본곡과 커버곡 쌍 80개로 이루어진 것으로 모두 160곡으로 구성되어 있으며 음원이 제공된다. 데이터셋 내의 임의의 음원을 질의하고 음악 유사도를 구한 후 가장 유사하다고 판정된 음악이 입력 음원의 커버곡이 맞을 경우의 확률인 P@1과 Mean of Average Precision(MAP)를 커버곡 검색 성능 지표로 사용하였다.

본 논문에서 사용한 음악 전곡 특징인 MOVE는 CREMA 특징[12]을 입력으로 사용하고 있다. Da-tacos 테스트 데이터셋의 경우 미리 매 11.6 ms 마다 CREMA 특징을 추출하여 제공하고 있다. covers80 데이터셋은 Reference [12]에서 제공되는 특징 추출 코드를 사용하여 음악 파일들을 모노로 바꾸고 44100 Hz로 샘플링 주파수를 맞춘 후 11.6 ms 마다 12차 CREMA 특

Table 1. Identification performance of the Da-tacos test dataset. Accuracy measures are precision at one, P@1, and the mean of average precision, MAP.

Feature type	Feature dim.	Music similarity	P@1	MAP
MOVE[7]	16000	Euclidean	0.737	0.507
MOVE-PCA[17]	256	Euclidean	n/a	0.507
Re-MOVE[17]	256	Euclidean	n/a	0.524
MOVE-BNNECK	256	Cosine	0.745	0.514
MOVE-BNNECK	256	PLDA	**0.762**	**0.535**

징을 직접 추출하였다. 얻어진 CREMA 특징을 시간축으로 1/8 로 다운 샘플링하여 93 ms 마다 CREMA 특징 벡터가 나오도록 하고 미리 학습된 MOVE 모델에 입력하여 전곡 특징을 구한다. 전곡 특징 간 유사도를 계산하여 커버곡 검색을 수행한다.

음악 유사도를 학습하기 위해서는 PLDA 확률 모델을 따로 학습해야한다. Da-tacos 데이터셋에는 테스트 데이터셋과 함께 분석 데이터셋이 제공된다. 분석 데이터셋은 커버곡쌍 5천개로 1만곡으로 구성되어 있다. 테스트 데이터셋과 같이 음원은 제공되지 않고 특징과 메타데이터만 제공된다. 분석 데이터셋으로부터 1만개의 MOVE 전곡 특징을 얻고 EM 알고리즘을 적용하여 Eq. (1)의 PLDA 모델 파라미터인 m, S, Σ를 학습하였다. PLDA 학습 시에 MSR identity toolbox의 코드를 변형하여 활용하였다. 본 실험에서는 Da-tacos 분석 데이터셋의 커버곡쌍에 최적화되도록 학습된 PLDA 모델을 활용하여 Eq. (3)의 로그 우도를 커버곡 검색을 위한 음악 유사도로 사용하였다.

Da-tacos 테스트 데이터셋에서 PLDA 기반 음악 유사도의 커버곡 검색 성능을 Table 1에 정리하였다. MOVE-BNNECK은 MOVE에 비해서 특징 차수가 낮지만 비슷한 검색 성능을 보이는 것을 확인할 수 있다. 또한 MOVE-BNNECK의 음악 유사도를 코사인 거리에서 Eq. (3)의 PLDA기반 로그 우도로 바꿀 때 기존 MOVE와 비교하여 5.5 % 정도 성능을 향상시킬 수 있음을 확인하였다. MOVE의 손실 함수를 바꾸어 재학습한 Re-MOVE 특징[17]의 성능 개선 정도인 3.4 %에 비해서 PLDA 기반 음악 유사도의 성능개선 정도가 더 크다. PLDA 기반 음악 유사도의 covers80 데이터셋에서 커버곡 검색 성능을 Table 2에 정리하였다. Da-tacos 테스트 데이터셋에서와 마찬가지로 MOVE와 MOVE-BNNECK의 성능은 비슷하였으며, PLDA 기반 음악 유사도가 MOVE-BNNECK의 성능을 개선하였다. 실험 결과로부터 제안한 PLDA기반 음악 유사도가 실험 대상 두 데이터셋 모두에서 커버곡 검색 성능을 향상시킬 수 있음을 확인하였다. 제안한 PLDA기반 음악유사도는 기존에 사용해온 Euclidean과 Cosine 거리와는 다르게 PLDA의 특징 변이 확률모델이 커버곡 변이 패턴을 학습하여 더 우수한 커버곡 검색 성능을 보인다. 추후 연구로 커버곡 생성 과정의 변이 별로 데이터셋을 구축하고 PLDA 잠재변수의 변이 별 강인성과 식별성에 대한 분석을 통해서 커버곡 검색성능을 더 개선할 수 있는 확률 모델을 찾을 수 있을 것으로 기대된다.

Table 2. Identification performance of the covers80 dataset. Accuracy measures are precision at one, P@1, and the mean of average precision, MAP.

Feature type	Feature dim.	Music similarity	P@1	MAP
MOVE	16000	Euclidean	0.819	0.846
MOVE-BNNECK	256	Cosine	0.819	0.844
MOVE-BNNECK	256	PLDA	**0.838**	**0.860**

IV. 결 론

커버곡 검색을 위한 음악 유사도를 PLDA를 이용하여 유도하였다. PLDA를 통해서 같은 커버곡 간의 특징 변이와 다른 음악들 간의 특징 변이를 각각 학습하여 확률 모델로 표현하고, 유사도 비교 대상 음악 특징들 간의 변이가 같은 커버곡 간의 변이에 속하는 지 다른 음악들 간의 변이에 더 가까운 지 우도를 계산하여 커버곡 관계 여부를 판정하였다. 두 커버곡 데이터셋에서 성능 비교 실험을 수행하여, PLDA 기반 음악 유사도가 커버곡 검색 성능을 향상시킬 수 있음을 보였다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원

의 2022년도 저작권기술 연구개발사업으로 수행되었음(과제명: 딥러닝을 활용한 고속 음악 탐색 기술 개발, 과제번호: CR202104004)

References

1. Y. V. S. Murthy and S. G. Koolagudi, “Content-based music information retrieval and its applications toward the music industry: A review,” ACM Comput. Surv. **51**, 1-46 (2019).
2. J. S. 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. F. Yesiler, G. Doras, R. M. Bittner, C. J. Tralie, and J. Serra, “Audio-based musical version identification: Elements and challenges,” IEEE Signal Process. Mag. **38**, 115-136 (2021).
4. 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).
5. J. S. Seo, “Cover song search based on magnitude and phase of the 2D Fourier transform” (in Korean), J. Acoust. Soc. Kr. **37**, 518-524 (2018).
6. G. Doras and G. Peeters, “Cover detection using dominant melody embeddings,” Proc. ISMIR, 107-114 (2019).
7. F. Yesiler, J. Serrà, and E. Gómez, “Accurate and scalable version identification using musically-motivated embeddings,” Proc. ICASSP, 21-25 (2020).
8. X. Du, Z. Yu, B. Zhu, X. Chen, and Z. Ma, “Bytecover: Cover song identification via multi-loss training,” Proc. ICASSP, 551-555 (2021).
9. S. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder, “Probabilistic models for inference about identity,” IEEE TPAMI, **34**, 144-157 (2012).
10. P. Rajan, A. Afanasyev, V Hautamäki, and T. Kinnunen, “From single to multiple enrollment i-vectors: Practical PLDA scoring variants for speaker verification,” Digit. Signal Process. **31**, 93-101 (2014).
11. D. Snyder, D. Garcia-Romero, G. Sell, A. McCree, D. Povey, and S. Khudanpur, “Speaker recognition for multi-speaker conversations using x-vectors,” Proc. ICASSP, 5796-5800 (2019).
12. B. McFee and J. P. Bello, “Structured training for large-vocabulary chord recognition,” Proc. ISMIR, 188-194 (2017).
13. A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” arXiv: 1703.07737 (2017).
14. H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, “Bag of tricks and a strong baseline for deep person re-identification,” Proc. CVPR workshops, 1487-1495 (2019).
15. F. Yesiler, C. Tralie, A. Correya, D. F. Silva, P. Tovstogan, E. Gómez, and X. Serrà, “Da-TACOS: A dataset for cover song identification and understanding,” Proc. ISMIR, 327-334 (2019).
16. *Covers80 Cover Song Data Set*, http://labrosa.ee.columbia.edu/projects/coversongs/covers80/, (Last viewed February 1, 2017).
17. F. Yesiler, J. Serrà, and E. Gómez, “Less is more: Faster and better music version identification with embedding distillation,” Proc. ISMIR, 884-892 (2020).

저자 약력

▸ 서 진 수 (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

▸ 김 정 현 (Junghyun Kim)

1999년 2월 : 전남대학교 전산학과 공학사
2001년 2월 : 전남대학교 전산학과 공학석사
2001년 3월 ~ 현재 : 한국전자통신연구원 책임연구원

▸ 김 혜 미 (Hyemi Kim)

2004년 2월 : 부산대학교 전자전기정보컴퓨터공학부 공학사
2006년 2월 : KAIST 전기 및 전자공학과 공학석사
2006년 2월 ~ 현재 : 한국전자통신연구원 선임연구원