

음악 장르 분류를 위한 부밴드 분해와 특징 차수 축소에 관한 연구

An investigation of subband decomposition and feature-dimension reduction for musical genre classification

서진수,^{1†} 김정현,² 박지현²

(Jin Soo Seo,^{1†} Junghyun Kim,² and Jihyun Park²)

¹강릉원주대학교 전자공학과, ²한국전자통신연구원 콘텐츠 연구본부
(Received December 13, 2016; accepted March 29, 2017)

초 록: 음악 장르는 음악 검색 및 분류 등의 정보 처리 시스템 구현에 있어서 필수적인 요소이다. 일반적으로 장르 분류를 위한 스펙트럼 특징은 음악의 화음 및 강약 구조를 표현하기 위해 부밴드로 분해하여 구해진다. 본 논문은 음악 장르 분류 성능 개선을 위한 특징 추출을 위한 부밴드 분해 방법에 관해 연구하였다. 또한 부밴드 음악 특징의 차수를 줄일 수 있는 방법에 대해서도 연구하였다. 널리 사용되고 있는 장르 데이터셋들에서 실험을 수행하여 널리 사용되고 있는 옥타브 스케일보다 세분화된 부밴드 분해가 장르 분류 성능을 향상시킬 수 있으며, 특징 차수 축소를 결합하여 분류기의 계산량도 줄일 수 있음을 보였다.

핵심용어: 음악 분류, 음악 장르, 부밴드 특징, 특징 차수 축소

ABSTRACT: Musical genre is indispensable in constructing music information retrieval system, such as music search and classification. In general, the spectral characteristics of a music signal are obtained based on a subband decomposition to represent the relative distribution of the harmonic and the non-harmonic components. In this paper, we investigate the subband decomposition parameters in extracting features, which improves musical genre classification accuracy. In addition, the linear projection methods are studied to reduce the resulting feature dimension. Experiments on the widely used music datasets confirmed that the subband decomposition finer than the widely-adopted octave scale is conducive in improving genre-classification accuracy and showed that the feature-dimension reduction is effective reducing a classifier's computational complexity.

Keywords: Music classification, Musical genre, Subband feature, Feature dimension reduction

PACS numbers: 43.75.Zz, 43.60.Uv

1. 서 론

음악 검색을 위해서 방대한 양의 음악 데이터가 존재하는 상황에서 빠르고 신뢰성 있게 정보를 제공해 줄 수 있는 오디오 신호 처리 기술의 중요성이 증대되고 있다.^[1] 음악을 표현하는 여러 요소들 가운데 일반적으로 사람의 음악 취향은 음악 장르로 표현되

는 경우가 많이 있다. 음악을 직접 듣고 수작업으로 장르를 분류하는 것은 많은 시간이 소요될 뿐만 아니라 인간의 실수에 의한 오분류의 가능성이 있다. 따라서 음악 신호의 특징을 이용하여 장르를 자동으로 분류하는 연구들이 큰 관심을 받아왔지만, 음악 장르라는 구분은 주관적인 요소가 있고 구분이 모호한 경우도 있으므로 여전히 어려운 문제로 남아 있다.^[2]

일반적으로 음악 장르 자동 분류는 Fig. 1과 같이 이루어진다.^[3,4] 준비 단계에서 분류기를 이미 장르를 알고 있는 음악의 특징 데이터를 이용하여 학습

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

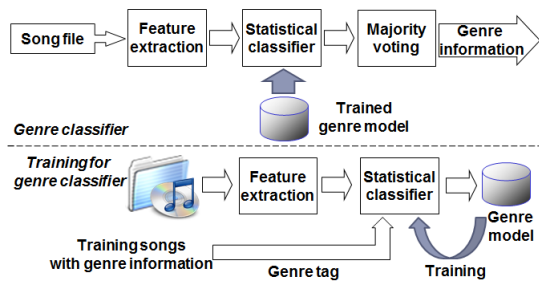


Fig. 1. Overview of the feature-based musical genre classification system.^[3]

한다. 학습된 분류기를 이용하여 입력 음악 파일의 특징에 대해서 장르 분류를 수행한다. 음악 장르 분류를 위해서는 장르 간의 차이를 두드러지게 할 수 있는 특징을 사용해야 하며, 일반적으로 스펙트럼 기반의 음색 특징이 널리 사용되어 왔다. 장르 분류에 사용되는 대표적인 스펙트럼 특징에는 MFCC (MelFrequency Cepstral Coefficients),^[2,3] OSC (Octave-based Spectral Contrast)^[5] 등이 있다. 이러한 스펙트럼 특징들은 음악 신호를 짧은 길이의 프레임(수 십에서 수 백 ms)으로 나누고 각 프레임에 푸리에 변환을 취하여 얻어진다. 프레임 레벨의 스펙트럼 특징만으로는 음악 장르를 신뢰성 있게 분류할 수 없으므로, 프레임 레벨 특징들을 모아서 1s에서 30s 길이의 음악 신호 세그먼트에서의 평균, 분산, 상관관계 등의 통계학적 특성을 구한다. 사용될 특징이 결정되면, 다음으로 그 특징에 맞는 분류기를 선택해야 한다. 일반적으로 계산량이 적은 선형분류기가 많이 쓰인다. 선형 분류기에는 SVM(Support Vector Machine), NB(Naïve Bayes), LDA(Linear Discriminant Analysis) 등이 있다. 이러한 선형 분류기는 학습을 통해서 오디오 특징을 장르 별로 잘 나눌 수 있는 선형(고차원 특징에 대해서는 실제적으로 초평면)의 경계를 찾아준다.

본 논문에서는 음악 장르 분류에 좋은 성능을 보인 스펙트럼의 극값에 대한 연구를 수행하였다. 이러한 극값은 화음의 구조와 세기 등의 정보와 밀접한 연관이 있어서 음악 장르 분류에 좋은 성능을 보였다.^[5] 기존의 스펙트럼 극값을 이용한 음악 분류 연구들에서는 특징을 추출할 부밴드의 대역폭과 개수 등의 파라미터들에 대한 장르 분류 성능 분석에

대한 연구가 없었다. 본 논문은 어떻게 음악 신호 스펙트럼을 나누어 분석하는 것이 장르 분류 성능에 도움이 되는 지에 대한 연구이다. 본 논문에서는 부밴드를 인간의 청각 지각을 모사한 멜 스케일^[6]로 나누고, 멜 스케일 상에서 부밴드의 개수를 가변해 가면서 부밴드 스펙트럼의 분포 특징을 추출하였다. 부밴드 스펙트럼 분포 특징으로는 기존 연구들^[3,5]에서 좋은 성능을 보인 스펙트럼의 최대값과 최소값인 극값을 고려하였다. 부밴드 개수의 증가에 따라서 늘어난 특징의 차수를 줄이기 위해서 LFDA(Local Fisher Discriminant Analysis)^[7,8]를 적용하였다.

본 논문은 음악 장르 분류를 위한 부밴드 선택 및 차원축소에 관한 연구이다. II장에서 부밴드 스펙트럼 극값 특징에 대해서 살펴보고, 구해진 특징에 LFDA 기반 특징 차원 축소 기법을 적용한다. III장에서 제안된 방법의 성능을 실험하고 결과를 비교 분석한다.

II. 멜 스케일 부밴드 스펙트럼 극값 및 LFDA 기반 특징 차원 축소

음악 분류기는 Fig. 1에 주어진 바와 같이 프레임 레벨 특징을 추출하고, 프레임들을 모은 세그먼트 레벨 특징 요약으로 구성되며, 세그먼트 레벨 요약 특징이 학습된 분류기의 입력으로 사용된다. 본 논문에서는 프레임 레벨 특징 추출 시 부밴드 개수와 세그먼트 레벨 특징의 차원 축소에 관해서 연구하였다.

2.1 부밴드 스펙트럼 극값 특징

부밴드 스펙트럼 극값과 그 차이값은 음악 장르 분류를 위해서 제안되었으며,^[5] 음성의 모음 인지도도 연관되어 있음이 알려져 있다.^[9] 음악 장르는 화음의 구조와 세기 등에 밀접한 관련이 있으므로, 기존의 MFCC처럼 부밴드 내에서 스펙트럼의 합을 구하는 것이 아니라 명시적으로 스펙트럼의 최대값과 최소값을 사용하는 것이 장르 분류 성능에 도움이 된다. 실제로 같은 특징 차수의 MFCC와 비교해서 스펙트럼 극값 특징이 더 나은 장르 분류 성능을 보였다.^[3]

음악 신호의 프레임에 FFT를 취해서 얻어진 스펙트럼의 b 번째 부밴드의 스펙트럼이 $\mathbf{P}_b = (P_{b,1}, P_{b,2}, \dots, P_{b,N_b})$ 로 주어진다 하자. 먼저 스펙트럼 \mathbf{P}_b 를 내림차순으로 정렬했을 때 $\mathbf{P}'_b = (P'_{b,1}, P'_{b,2}, \dots, P'_{b,N_b})$ 라고 하면, b 번째 부밴드의 최대값 $peak_b$ 와 최소값 $valley_b$ 값은 주변평활 계수인 α 값에 따라서 아래와 같이 주어진다(본 논문에서는 $\alpha = 0.2$ 로 사용함).^[5]

$$\begin{aligned} peak_b &= \log\left(\frac{1}{\alpha N_b} \sum_{i=1}^{\alpha N_b} P'_{b,i}\right), \\ valley_b &= \log\left(\frac{1}{\alpha N_b} \sum_{i=1}^{\alpha N_b} P'_{b,N_b-i+1}\right). \end{aligned} \quad (1)$$

기존의 스펙트럼 부밴드 특징을 사용한 연구들은 스펙트럼을 옥타브 스케일의 부밴드로 나누었다. 일반적으로 각 옥타브 스케일의 부밴드는 겹치지 않으며 시작과 끝 주파수는 Hz 단위로 0~200, 200~400, 400~800, 800~1600, 1600~3200, 3200~6400으로 총 6개의 주파수 대역이 사용되었다.^[5] 하지만 이러한 옥타브 스케일 부밴드 사용에 대해서 어떠한 성능 검증도 이루어지지 않았다. 본 논문에서는 특징 추출 시에 인간의 청각을 모사한 멜 스케일 상에서 부밴드의 개수를 가변하였다. 인간의 청각에 대한 실험을 통해서, 스펙트럼의 주파수 1000 Hz 이하에서는 주파수와 멜스케일이 선형 관계에 가깝고, 그 이상에서는 로그 함수 형태에 가깝다고 알려져 있다.^[10] 다양한 변환식이 존재하지만, 본 논문에서 사용한 주파수(f)와 멜(m) 스케일 변환 식은 다음과 같다.^[6]

$$m = 1127 \log_e(1 + f/700). \quad (2)$$

위 변환식을 사용하여 0에서 6400 Hz 주파수 대역을 멜 스케일로 나타내면 Fig. 2와 같다. 부밴드 극값 특징을 추출하기 위해서 멜 스케일 상에서 균등하게 대역을 분해하였다. 만약 4개의 대역을 사용한다면 멜 스케일 상에서는 0에서 2611사이를 균등 분할하여 0~653, 653~1306, 1306~1958, 1958~2611이며, 이는 주파수 대역에서 0~551, 551~1528, 1528~3281, 3281~6400에 해당한다. 본 논문에서는 인간 청각 모

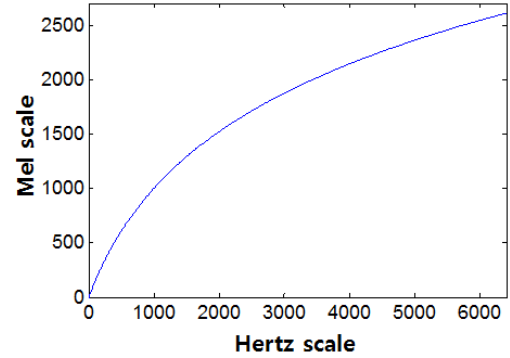


Fig. 2. Frequency versus mel scale.

델에 기반한 멜 스케일 상에서 부밴드를 분해하고 몇 개의 부밴드를 사용하는 것이 음악 장르 분류에 적합한 지를 확인하였다.

2.2 세그먼트 기반 특성 추출 및 LFDA기반 차원 축소

음악 프레임에서 얻은 부밴드 특징을 세그먼트 레벨에서 요약하여 장르 분류기에 사용한다. 세그먼트의 길이는 1s에서 30s 사이로 다양한다. 본 논문에서는 3s를 사용하였다. 각 세그먼트의 요약 특징으로 장르 분류를 수행하고, 각 세그먼트의 분류 결과를 전곡에 대해서 취합하여 최종 장르 결정을 한다. 일반적으로 음악 파일 내의 모든 세그먼트들의 분류 결과들을 모아서 최다수 선택 원칙을 적용한다.

세그먼트 요약 특징으로는 평균과 표준편차가 널리 사용된다. Eq. (1)에 주어진 바와 같이 음악 신호 각 프레임에서 얻어진 스펙트럼을 B 개의 부밴드로 나누어 최대값과 최소값을 구하면, 프레임 별로 $2B$ 개의 특징이 구해진다. 인접한 M 개의 프레임인 세그먼트에 대해서 각 프레임 특징을 평균과 표준편차로 요약하면 최종 요약 특징은 $4B$ 차수를 가지게 된다. 부밴드 개수 B 가 커짐에 따라서 특징 차수도 늘어나므로 이를 줄일 수 있는 차원 축소 방법이 필요하다. 본 논문에서는 LFDA를 적용하였다.^[7] LFDA를 통해서 특징 분포의 국소 이웃 구조를 보존하고 특징 차원을 축소하면서도 장르 차이 정보를 최대화할 수 있는 변환 행렬을 구할 수 있다. LFDA 변환 행렬 학습 과정은 다음과 같다. 학습에 사용할 세그먼트의 개수가 n 개이고 분류대상 장르의 개수가 c 일 때, 각

세그먼트에서 얻은 $4B$ 차원 특징 벡터를 $X_i (i=1, 2, \dots, n)$ 라고 하고 각 특징 벡터가 속하는 음악장르 레이블을 $y_i \in (1, 2, \dots, c)$ 라고 하자. 같은 장르로부터 얻은 특징 벡터간의 산포도 행렬과 다른 장르에서 얻은 특징 벡터간의 산포도 행렬을 특징 벡터간의 인접 행렬 A 를 가중치로 하여 구한다. 본 논문에서 인접 행렬 A 는 n 행 n 열 행렬로, A_{ij} 는 X_i 와 X_j 간의 코사인 거리로 한다.

$$S^W = \frac{1}{2} \sum_{i,j=1}^n A_{i,j}^W (X_i - X_j)(X_i - X_j)',$$

$$S^B = \frac{1}{2} \sum_{i,j=1}^n A_{i,j}^B (X_i - X_j)(X_i - X_j)'. \quad (3)$$

Eq. (3)에서 가중치 행렬들은 각 장르에 속한 학습 특징의 개수가 n_c 개라고 할 때 다음과 같이 주어진다.

$$A_{i,j}^W = \begin{cases} A_{i,j}/n_c & \text{if } y_i = y_j = c \\ 0 & \text{if } y_i \neq y_j \end{cases},$$

$$A_{i,j}^B = \begin{cases} A_{i,j}(1/n - 1/n_c) & \text{if } y_i = y_j = c \\ 1/n & \text{if } y_i \neq y_j \end{cases}. \quad (4)$$

얻어진 산포도 행렬들로부터 장르 차이가 두드러질 수 있도록 변환행렬 T_{LF} 는 다음과 같이 정의된다.^[7]

$$T_{LF} = \arg \max_T [(T^T S^W T)^{-1} T^T S^B T]. \quad (5)$$

Eq. (5)는 행렬의 고유값과 고유벡터를 구하는 문제인 $S^B \psi = \lambda S^W \psi$ 로 풀 수 있다. 얻어진 고유벡터를 고유값의 크기에 따라 내림차순으로 r 개를 취하여, $4B$ 행 r 열 변환행렬 T_{LF} 는 다음과 같이 주어진다.

$$T_{LF} = (\sqrt{\lambda_1} \psi_1, \sqrt{\lambda_2} \psi_2, \dots, \sqrt{\lambda_r} \psi_r). \quad (6)$$

최종적으로 $4B$ 차원 특징벡터 X_i 에 변환행렬 T_{LF} 를 곱해서 차원 축소된 특징벡터 Z_i 를 구한다.

$$Z_i = T_{LF} X_i. \quad (7)$$

얻어진 특징벡터 Z_i 를 이용하여 장르 분류기를 학

습한다. 차원 축소를 통해서 분류기 학습 및 실제 적용 시에 $4B$ 차원이 아닌 r 차원 벡터를 사용함으로써 계산량을 줄일 수 있다.

III. 실험 결과

본 장에서는 II장에서 제시한 바와 같이 부밴드 스펙트럼 극값 기반 음악 장르 분류기에 부밴드 개수를 가변 시켜가면서 장르 분류 성능을 확인하였다. 장르 분류 성능 검증을 위해서 널리 사용되고 있는 두 가지 음악 데이터셋인 ISMIR과 GTZAN을 사용하였다. ISMIR 데이터셋은 클래식, 일렉, 재즈, 메탈, 락, 세계 음악 이렇게 6가지 장르로 구분되어있으며 1458곡으로 이루어져 있다. 전체 곡중 절반인 729곡은 학습으로 나머지 729곡은 분류 성능 검증용으로 제공된다. GTZAN 데이터셋^[11]은 블루스, 클래식, 컨츄리, 디스코, 힙합, 재즈, 메탈, 팝, 레게, 락의 10개의 장르에 각각 100곡씩 30s 길이의 1000개의 음악파일로 이루어져 있다. 주어진 1000개의 음악 파일을 500곡씩 학습과 테스트셋으로 나누었다. 같은 가수의 노래들이 학습과 테스트셋 양쪽에 나누어 포함될 경우 분류 성능이 높게 나오는 경향이 있다. 이는 특정 가수의 음색 유사성이 분류기에 영향을 미치기 때문으로, 공정한 장르 분류 알고리즘 성능 평가가 되지 못한다. 따라서 본 논문에서는^[12]에서 제안한 바와 같이 공정한 장르 분류 성능 평가가 되도록 특정 가수의 노래들이 학습과 테스트셋 중에서 한쪽에만 포함되도록 조정하였다.

실험에 사용되는 음악 파일들을 모노로 바꾸고 22050Hz로 샘플링 주파수를 맞춘 후, 1024 길이의 해닝 윈도우를 50%씩 겹쳐 가면서 적용하고 FFT를 취해서 각 프레임의 스펙트럼을 구한다. 각 프레임의 스펙트럼을 멜스케일 상에서 균등하게 부밴드로 나누고 각 부밴드로부터 Eq. (3)의 최대와 최소값을 구한다. 3s 길이에 해당하는 인접한 128개의 프레임들을 모아서 세그먼트를 구성하고 각 부밴드 특징의 평균과 표준편차를 구해서 세그먼트 특징으로 이용한다. 세그먼트 특징들로 선형 SVM 장르 분류기를 학습하고 분류 실험을 수행하였다.

부밴드의 개수에 따른 분류 성능은 Figs. 3과 4에

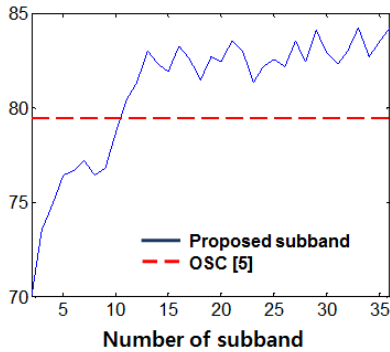


Fig. 3. Classification accuracy (%) versus number of mel subband for ISMIR genre dataset.

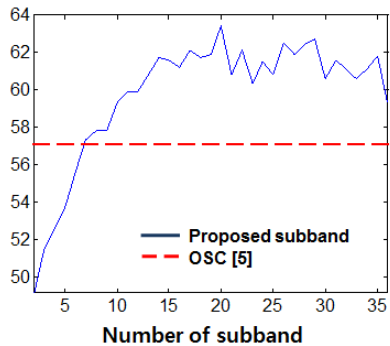


Fig. 4. Classification accuracy (%) versus number of mel subband for GTZAN genre dataset with singer filter.

도시하였다. 실험 결과 부밴드 개수가 늘어남에 따라 특징 차수가 높아지고 분류성공률도 높아짐을 확인할 수 있다. 다만 부밴드 개수가 14개 이상이 되면 더 이상 성능이 크게 증가하지 않음을 알 수 있다. 기존의 OSC 결과^[5]와 비교할 때 더 많은 수의 부밴드를 사용하는 것이 장르 분류 성능을 향상시킬 수 있음을 알 수 있다. 기존의 6개 옥타브 스케일 밴드를 사용한 OSC결과와 비교하여 ISMIR과 GTZAN 데이터셋 모두에서 최대 5% 내외의 분류 성공률 개선이 가능함을 확인하였다.

부밴드 개수를 증가시키에 따라 분류기에 사용되는 특징의 차수도 증가되는 문제가 있다. 부밴드 개수가 16개인 경우 ($B=16$)에 대해서 LFDA를 이용하여 특징의 차수를 축소하였으며, 특징 차수에 따른 분류 성능을 Figs. 5와 6에 도시하였다. 실험에서는 LFDA 외에도 PCA(Principal Component Analysis), DCT (Discrete Cosine Transform), Haar 변환도 적용하여 성

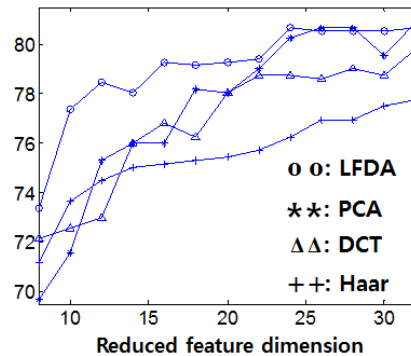


Fig. 5. Classification accuracy (%) versus feature dimension for ISMIR genre dataset.

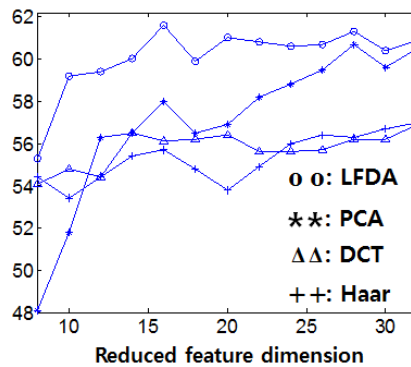


Fig. 6. Classification accuracy (%) versus feature dimension for GTZAN genre dataset with singer filter.

능을 비교하였다. PCA는 학습데이터로부터 얻은 공분산 행렬의 고유값과 고유행렬로부터 변환을 구한다. LFDA와는 달리 PCA에서는 학습데이터의 장르 레이블은 고려하지 않는다. DCT와 Haar 변환은 고정된 베이스를 가지며, 본 논문에서 사용된 DCT와 Haar 변환의 파라미터는 Reference [13]에 나온 것과 같다. 특징 차수 축소를 많이 할 경우 LFDA의 성능이 우수하였다. 이는 LFDA가 장르별 특징의 분포 특성을 반영하여 변환 후 같은 장르에서 얻은 특징은 가깝게하고 다른 장르에서 얻은 특징은 멀어지도록 변환 행렬을 구하기 때문으로 생각된다. 두 데이터셋 모두에서 특징의 분포적 특성을 고려하는 LFDA와 PCA가 고정 베이스를 사용하는 DCT와 Haar 변환에 비해서 우수한 성능을 보였다.

기존의 OSC 방법^[5]과 같은 차수인 24차로 줄인 경우의 분류 성공률은 Table 1에 주어졌다. 같은 차수로 줄일 경우에도 OSC와 대비해서 각 데이터셋 별

Table 1. Comparison of classification accuracy (%) with the conventional features, MFCC and OSC. For all the following cases, the number of segment-level features was 24. In case of LFDA, PCA, DCT, we used 16 subbands ($B = 16$) in extracting features and reduced feature dimension to 24.

Method/dataset	ISMIR dataset	GTZAN dataset
MFCC	76.68	53.5
OSC	79.42	57.1
LFDA Projection	80.66	60.6
PCA Projection	80.25	58.8
DCT Projection	78.74	55.6

로 1.2%와 3.5% 분류 성능을 개선함을 확인하였다. MFCC와 대비해서는 각 데이터셋 별로 4%와 7% 분류 성능을 개선함을 확인하였다. 이러한 실험 결과들로부터 오디오 스펙트럼 분석 시에 부밴드의 위치 및 개수를 정하는 것이 성능에 중요한 영향을 미치며, 특징 분포 및 장르 분류에 적응적인 변환행렬을 이용하여 특징 차수를 줄일 수 있음을 확인하였다.

IV. 결 론

본 논문은 음악 장르 분류를 위한 부밴드 선택 방법에 관한 논문이다. 음악 장르 분류를 위한 부밴드 특징의 성능을 분석하여 최적의 부밴드 개수를 찾았다. 또한 부밴드 음악 특징의 차수를 줄일 수 있도록 특징 분포 및 장르 분류에 적응적인 변환행렬을 적용하는 방법에 대해서도 연구하였다. 널리 사용되고 있는 장르 데이터셋들에서 실험을 수행하여 세분화된 부밴드 분해가 장르 분류 성능을 향상시킬 수 있으며, 특징 차수 축소를 결합하여 분류기의 계산량도 줄일 수 있음을 보였다.

감사의 글

본 연구는 문화체육관광부 및 한국저작권위원회의 2016년도 저작권기술개발사업의 연구결과로 수행되었음(연관 저작물 검색과 저작권 침해 예방을 위한 예측 탐지 기술 개발).

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. B. L. Sturm, "A survey of evaluation in music genre recognition," in *Proc. AMR-2012*, 29-66 (2012).
3. J. Seo, "Study on the performance of spectral contrast MFCC for musical genre classification" (in Korean), *J. Acoust. Soc. Kr.* **29**, 265-269 (2010).
4. S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, "Music genre classification system using decorrelated filter bank" (in Korean), *J. Acoust. Soc. Kr.* **30**, 100-106 (2011).
5. D. Jiang, L. Lu, H. Zhang, J. Tao, and L. Cai, "Music type classification by spectral contrast feature," in *Proc. ICME-2002*, 113-116 (2002).
6. D. O'Shaughnessy, *Speech Communication: Human and Machine* (Addison-Wesley, Boston, 1987), pp. 150.
7. M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.* **8**, 1027-1061 (2007).
8. S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *Proc. CVPR*, 3318-3325 (2013).
9. P. Loizou and O. Poroy, "Minimum spectral contrast needed for vowel identification by normal-hearing and cochlear implant listeners," *J. Acoust. Soc. Am.* **110**, 1619-1627 (2001).
10. S. Umesh, L. Cohen, and D. Nelson, "Fitting the mel scale," in *Proc. ICASSP*, 217-220 (1999).
11. G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech and Audio Process.* **10**, 293-302 (2002).
12. B. L. Sturm, "An analysis of the GTZAN music genre dataset," in *Proc. ACM workshop on Music information retrieval with user-centered and multimodal strategies*, 7-12 (2012).
13. A.K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, Trenton, 1989), pp. 150-161.

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: 한국전자통신
연구원 선임연구원
2008년 3월 ~ 현재: 강릉원주대학교 전자
공학과 조교수, 부교수

▶ 김정현 (Junghyun Kim)



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

▶ 박지현 (Jihyun Park)



1997년 2월: 서강대학교 전산학과 공학사
1999년 2월: 서강대학교 컴퓨터학과 공학
석사
2010년 2월: 충남대학교 컴퓨터공학과
공학박사
1999년 5월 ~ 현재: 한국전자통신연구원
책임연구원