음악 무드 분류에서 음향 특성과 차원 감쇄 기법의 효과 연구*

한병준[1] 노승민[2] 황인준[1]

[1] 고려대학교 전기전자전파공학부

{hbj1147, ehwang04}@korea.ac.kr

[2] 아주대학교 정보통신전문대학원

anycall@ajou.ac.kr

Effects of Acoustic Features and Dimensionality Reduction Techniques in Musical Mood Classification

Byeong-jun Han[1] Seungmin Rho[2] Eenjun Hwang[1]

[1] School of Electrical Engineering, Korea University

[2] Graduate School of Information and Communication, Ajou University

요 약

인터넷을 비롯한 통신 네트워크의 발전으로 개개인의 콘텐츠 수요가 증가함에 따라 다양한 콘텐츠 욕구를 충족시키기 위한 추천 시스템이 대두되고 있으며, 이러한 추천 시스템의 기반 기술로써 내용 기반 검색 기술의 필요가 증가하고 있다. 본 논문에서는 대표적인 멀티미디어 콘텐츠의 하나인 음악의 무드를 내용 기반으로 분류하기 위해, 음악 비트 검출에 기반한 프레임화를 적용하였으며, 스펙트럼의 고조파를 좀더 강조하기 위한 HDS (Harmonic Distribution Spectrum)을 제안하였다. 또한 다양한 차원 감쇄 기법과 분류기를 이용한 실험을 통해 무드 분류 시스템의 성능 비교를 진행하였다.

1. 서 론

컴퓨터 기반의 통신 네트워크가 생활과 밀착되면서 사용자들은 다양하고 수많은 멀티미디어 콘텐츠를 쉽게 접할 수 있게 되었고 이로 인하여 멀티미디어 콘텐츠 산업이 크게 성장하였다. 특히 멀티미디어 콘텐츠의 정형화된 형태로써 음악이나 동영상 등의 판매가 급증하였으며, CD나 DVD와 같이 기존 물리적 미디어 형태로 배포되던 산업으로부터 MP3, VoD 스트리밍 등과 같은 네트워크 기반 미디어 배포 산업으로 멀티미디어 콘텐츠 산업의 중심축이 이동하게 되었다.

음악 콘텐츠 산업의 큰 변화는 MP3와 같은 멀티미디어 데이터 파일 형태로 콘텐츠가 포장되고, 네트워크를 통한 판매가 이루어지는 것이다. 따라서 사용자는 이전보다 더 적은 부피의 미디어에 더 많은 콘텐츠를 소유할 수 있게 되었다. 그러나 개개인이 많은 콘텐츠를 소유하게 됨에 따라, 콘텐츠를 관리하기 위한 소프트웨어의 필요성이 증가하게 되었다. 또한, 이러한 소프트웨어의 필수 기능으로써, 사용자가 소유한 수많은 콘텐츠 중 상황에 맞는 콘텐츠를 찾아 추천할 수 있는 기능이 필요하게 되었다. 비록 Microsoft사의 Windows Media Player, Apple사의 iTunes 등이 이러한 필요를 부분적으로 만족시켜 주고는 있지만, 아직까지는 사용자 또는 콘텐츠 배포 업체로부터 입력한 주석 기반 검색 (Annotation-based Retrieval)이 구현 및 상용화 되어 있을 뿐이다.

텍스트 주석에 기반한 검색은 텍스트 주석의 내용이 완벽하다는 가정 하에서 검색이 가능해진다. 그러나 최근에는 음악의 무드, 장르 등이 복합화되고 다양화됨에 따라, 기존의 텍스트 주석만으로 음악 콘텐츠를 완벽하게 표현할 수 없게 되었다. 따라서 음악 정보 검색 (MIR; Music Information Retrieval) 분야에서도 내용 기반 검색 (Content-Based Retrieval)을 위한 기술 연구가 필요하게 되었다.

이를 위해 본 논문에서는 음악 콘텐츠를 표현하는 특징 중 음악적 분위기를 표현하는 대표적 특징인 무드 (mood) 정보를 분류하는 스키마의 비교 연구하고 실험을 통하여 그들의 성능을 비교하고자 한다. 본 논문의 구성은 다음과 같다. 2절에서는 기존의 음악 무드 및 장르 검출을 위한 다양한 연구를 알아본다. 3절에서는 음악의 주파수 특성을 추출하기 위한 기존의 STFT (Short-Term Fourier Transform)을 확장하여, 음성의 고조파 특성을 반영한 HDS (Harmonics Distribution Spectrum)을 제안하고, 이를 기반으로 다양한 음성 특성을 추출한다. 4절에서는 PCA, NMF 등의 다양한 차원 감쇄 기법과, NN (신경망 네트워크), GMM (가우시안 혼합 모델), 그리고 SVM 등의 분류기를

* 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2007-313-D00758)

간략하게 소개한다. 마지막으로 5절에서는 분류 스키마를 기반으로 무드 분류 성능을 측정한 결과를 보인다.

2. 관련 연구

음악의 특성을 분류하기 위한 연구가 다양하게 진행되어 오고 있다. 매 해마다 전세계 음악 정보 검색 연구의 정량화된 평가를 진행하는 MIREX[1]는 음악의 장르, 무드와 같은 거시적 특성뿐만 아니라 음악 구조 분석, onset detection, QBSH (Query-By-Singing/Humming) 등과 같은 다양한 콘테스트를 진행하고 있다.

많은 음악 무드 분류 연구자들은 Thayer의 심리 모델을 기반으로 연구를 진행하고 있다. Thayer는 활동도(arousal)와 압력(valence)에 따른 A/V model을 제안하였다. 그는 A/V model에 사람의 감정을 나타내는 형용사 단어들을 수학적으로 배치할 수 있다고 생각하였으며, Lu et al.[2]은 그들의 연구에서 차원 감쇄 기법 및 kNN, GMM 등의 분류기를 사용하여, 이러한 배치의 가능성이 있음을 보였다.

3. 특성 벡터 추출

이 절에서는 음악의 내용 특성을 추출하기 위해 접근한 방법에 대해 설명한다. 우선, 음악 신호의 프레임화 방법으로 비트 (beat) 기반의 framing을 진행하였다. 이때, 비트 검출 (beat detection)을 위해 Cover Songs[3]의 비트 추적 엔진을 사용하였다. 이후, 기존의 STFT (Short-Term Fourier Transform)의 주파수 도메인 결과 및 이의 고조파 특성을 부각시킨 새로운 주파수 도메인 계산 방법 인 HDS (Harmonic Distribution Spectrum) 의 결과를 추출하였다. 이를 기반으로 주파수 도메인의 음향 특성인 MFCC, SS, SC, SF, 그리고 시간 도메인의 음향 특성인 AE, ZCR 등을 추출하였다.

3.1. 비트 검출 기반의 프레임화

기존의 오디오 신호 처리를 위한 프레임화 방법은 입력 신호 시퀀스를 상황에 따라 10ms~1s 길이의 시퀀스로 분리하여 이에 시간 및 주파수 기반 음향 특성을 추출하는 방법이 일반적으로 적용되고 있다. 그러나 이러한 접근 방법은 음악적 템포가 지나치게 느리거나 지나치게 빠르기 때문에 정해진 시퀀스 길이 내에 표현되는 음향 특성의 밀도가 균일하지 않을 때 문제가 발생한다.

따라서 본 연구에서는 입력되는 음악 신호의 비트 (beat)를 검출하고 이를 프레임화의 기본 단위의 하나로 활용하는 것을 제안한다. 여기서 비트란 음악적 리듬을 나타내는 기본 박자의 길이를 나타낸다. 가령, 4/4 박자를 가지는 음악의 경우, 1/4 분 음표가 해당 음악의 기본 비트 단위가 된다.

본 연구에서는 다양한 기본 비트 검출을 위한 알고리즘 중 D. Ellis 및 G. Poliner에 의해 제안 된 Cover Songs[3]의 비트 추적 알고리즘을 사용한다. 본 알고리즘은 음악 정보 검색 분야의 저명 콘테스트인 2006년 MIREX 콘테스트 [1]에서 우수한 성능을 보였다. 사용한 알고리즘은 250Hz 주파수 샘플링에서 onset 강도를 계산하고, 이의 1차 차분과 적절한 경계치를 이용하여 비트 후보를 검출한다. 이후, 다이내믹 프로그래밍 방법을 사용하여 최적화된 비트 시퀀스를 계산한다.

그러나, 음악의 템포 (tempo)가 약 60~200 정도의 BPM (Beats-Per-Minute) 을 가지는 점을 감안하면, 한 비트에 해당하는 시퀀스의 길이는 1초 이상으로 상당히 긴 편이다. 따라서 본 연구에서는 하나의 비트를 8개의 균일한 시퀀스로 나누어, 이를 하나의 프레임으로 정의한다. 또한, 노래의 첫 부분 혹은 끝 부분의 경우, 비트 검출이 어려우므로 음악의 전체 템포를 이용하여 비트를 가정함으로써 프레임을 추출한다.

3.2. HDS (Harmonic Distribution Spectrum)

일반적으로 음악은 고조파에 의한 조화 (harmonic) 특성을 가지는 피아노, 바이올린, 사람 목소리 등의 조화 악기와 부조화 (inharmonic) 특성을 가지는 실로폰, 드럼 등과 같은 악기의 합동 연주로 구성된다. 이 때 조화 특성을 가지는 악기의 주파수 도메인 분석에서 일정한 간격을 두고 나타나는 다수의 고조파가 peak를 표현하고 있음을 알 수 있다. 이 때 이러한 고조파는 기본 고조파와 배수 고조파 (overtones)들로 구성되어 있다[4]. 또한, 일반적으로 부조화 악기는 음악의 리듬 (rhythm) 특성을 주도한다[5]. 따라서 음악의 분석을 좀더 효율적으로 하기 위해서는 조화 특성이 반영된 악기 신호를 중심으로 분석할 필요가 있다.

제안하는 HDS (Harmonic Distribution Spectrum)은 이러한 의도가 반영된 방법으로, 기존의 STFT 기반의 스펙트로그램 (spectrogram) 에 비해 고조파 특성을 더욱 부각시켜주는 보완된 스펙트럼 분석 방법이다. 또한 프레임화로 인해 나타나는 불연결성 (discontinuity) 에 의한 주파수 도메인의 에너지 누수 (spectral leaking)를 각 고조파에 집중시키기 위해, 고조파의 배음 주파수에 따라 동적인 길이를 가지는 윈도우 함수를 곱셈 적용한다.

우선, 각 프레임에 대해 STFT를 적용하여 스펙트럼을 계산한다. 이 때, 프레임의 길이가 서로 다르므로, 전체 스펙트럼의 크기를 프레임 길이로 나누어주는 에너지 표준화 작업이 필요하다. 이는 다음의 수식 (1)과 같이 나타낼 수 있다.

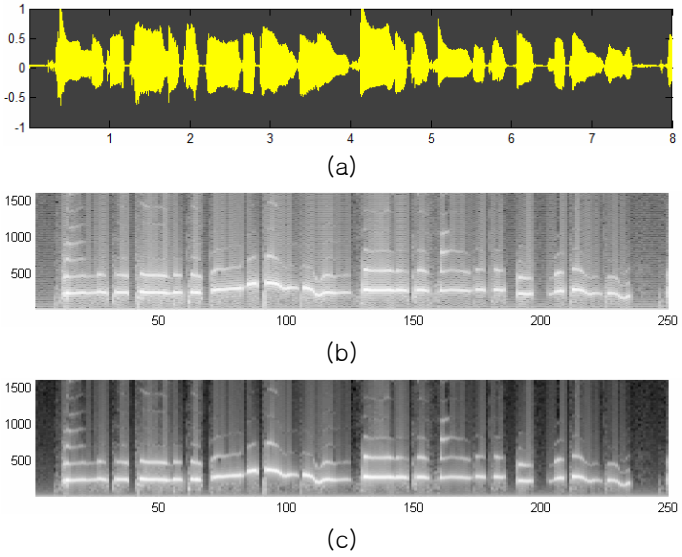

그림 1 (a) 사람의 목소리 신호 (b) STFT에 의한 스펙트로그램 결과 (c) HDS에 의한 스펙트로그램 결과

$$\bar{F}(x, f) = \frac{1}{N}\sum_{k=1}^{N} x(k)\cdot e^{\frac{-jk2\pi f}{f_s}} \quad (1)$$

이 때, x는 프레임화 된 시퀀스이며, N은 시퀀스의 길이이다. 또, f_s는 샘플링 주파수이므로, (1)의 스펙트럼은 $f=1,2,\ldots,f_s$의 해상력을 갖는다.

HDS는 기본 고조파가 해당 고조파로부터 파생되는 배수 고조파(overtones)의 에너지를 가능한 한 많이 포함한다고 가정한다. 따라서 에너지 누수를 고려하지 않은 HDS는 다음의 수식 (2)와 같이 정의될 수 있다.

$$\mathrm{HDS}_{\mathrm{incomplete}}(x,f) = \sum_{m=1}^{M} \bar{F}(x, mf) \quad (2)$$

이 때, 수식 (2)의 M은 기본 고조파에 포함하기 위한 최대 배수 고조파의 개수이다. 또한 f는 기본 고조파의 주파수인데, 실제 입력 신호에서 기본 고조파의 주파수가 아니더라도 스펙트럼에서의 일관성을 나타내기 위해 기본 고조파라고 가정한다.

한편, 에너지 누수를 감안하여 기본 고조파 및 배수 고조파의 거리 및 길이에 따라 아래의 수식 (3)과 같은 동적 길이를 가지는 다양한 윈도우 함수를 적용하여 HDS를 구할 수 있다.

$$\mathrm{HDS}(x, f) = \sum_{m=1}^{M}\sum_{l=-md}^{l=md} \bar{F}(x, mf + l)\cdot w_{2md+1}(l + md) \quad (3)$$

이 때, 수식 (3)의 d는 기본 고조파에 대해 에너지 누수를 계산하기 위한 최대 거리이다. 또한, w_k는 k 길이를 가지는 윈도우 시퀀스를 나타낸다.

그림 1은 사람의 목소리 신호(그림 1(a))에 대해 일정 구간으로 프레임화를 진행하고, 각 프레임에 대해 STFT(그림 1(b)) 및 HDS(그림 1(c))를 적용한 후 그 결과를 스펙트로그램으로 나타낸 것이다. 기존의 STFT 기반 스펙트로그램은 주변의 에너지 대비 고조파 에너지 차이가 크지 않은 반면, HDS에 기반한 스펙트로그램은 고조파와 주변 에너지의 대비가 뚜렷한 것을 볼 수 있다. 따라서, HDS는 에너지 누수에 의한 고조파 검출 문제를 어느 정도 극복하기 위한 기반 스펙트럼 분석 방법임을 알 수 있다.

표 1 음향 특성 및 차원 정리

음향 특성	특징	차원
SC	해당 프레임의 밝기 및 중심주파수	1
SR	스펙트럼의 모양 정량화	1
SF	스펙트럼 변화량 측정	1
MFCC	Mel-주파수 영역에서의 계수	12
AE	단위 프레임의 에너지 정량화	1
ZCR	단위 프레임의 시퀀스 영교차율	1
	통계적 특성 고려 (AE 제외)	(x 3)
합계	(1+1+1+12+1) x 3 + 1 =	**49**

3.3. 음향 특성 추출

기존의 분류 연구[2][6]에서 음향 특성을 추출하기 위한 다양한 방법이 제안되어 있다. 본 연구에서는 HDS 기반 분석에서 기존의 음향 특성을 적용하였을 때 정확도 성능 변화를 관찰하는 것이 목적이다. 따라서 표와 같이 기존에 사용된 다양한 대표적인 방법들을 응용하여 시너지 효과를 내는 것에 주력하였다.

STFT 및 HDS와 같은 주파수 기반 환경에서 음향 특성으로는 통계적 분석에 기반한 SC (Spectral Centroid), SR (Spectral Rolloff), SF (Spectral Flux)와 같은 방법이 대표적이다. 이들은 각각 다음 수식 (4)-(6) 과 같이 정의된다.

$$\mathrm{SC}(x) = \frac{\sum_{f=1}^{f_s} f\cdot \bar{F}(x, f)}{\sum_{f=1}^{f_s} \bar{F}(x, f)} \quad (4)$$

$$\sum_{f=1}^{\mathrm{SR}(x)} \bar{F}(x, f) = 0.85\cdot \sum_{f=1}^{f_s} \bar{F}(x, f) \quad (5)$$

$$\mathrm{SF}(x) = \sum_{f=1}^{f_s} \left\{\bar{F}(x_i, f) - \bar{F}(x_{i-1}, f)\right\}^2 \quad (6)$$

SC (Spectral Centroid)는 주파수 스펙트럼의 중심(centroid)를 구하는 것으로, 해당 시퀀스의 주파수 밝기 (brightness)를 측정하는 데에 사용된다. 따라서 이 음향 특성을 사용하면 전반적으로 어느 주파수 대역에 음이 집중되는 지를 계산할 수 있다. 일반적으로 높은 옥타브의 음이 많이 몰린 시퀀스의 경우 이 값이 높게 나타난다.

SR (Spectral Rolloff)는 전체 스펙트럼 에너지 합의 85%에 해당하는 주파수 범위를 말한다. 따라서 수식 (5)에서와 같이, 적절한 주파수 영역을 찾는 것이 SR을

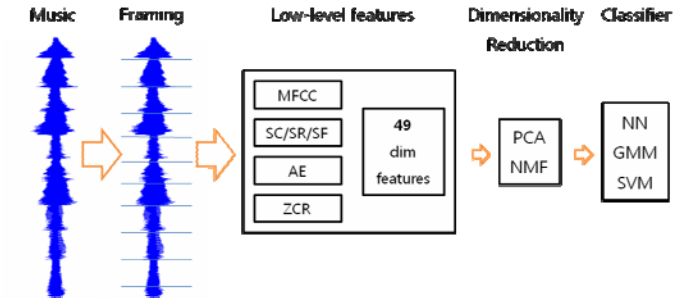

그림 2 음악 무드 분류를 위한 스키마

구하는 것이다. SR은 많이 사용되지 않는 주파수 영역을 제외한, 주요 주파수 영역을 구하는 데에 사용되며, 스펙트럼의 모양을 정량화하는 데에도 사용된다.

마지막으로, SF (Spectral Flux)는 현재 프레임과 이전 프레임 간의 스펙트럼 에너지 변화의 제곱 합이다. 이 음향 특성은 이전 프레임과 현재 프레임 간에 스펙트럼 변화가 얼마나 있는 지 측정하기 위한 방법이다.

한편, MFCC (Mel-Frequency Cepstral Coefficients)는 사람 귀에 친숙한 Mel-주파수 영역으로의 변환을 통해 특정 영역의 주파수 분포를 계수화하는 방법으로, 음성 인식 등의 신호 처리 분야에서 널리 쓰이고 있는 음향 특성이다. 일반적으로 첫번째 계수를 제외한 2~13번째 계수를 이용한 실험이 많으며, MFCC를 계산하기 위한 방법은 [7]에 자세히 언급되어 있다.

주파수뿐만 아니라 시간 기반 음향 특성도 널리 이용되고 있다. AE (Average Energy)는 단위 프레임의 평균 에너지를 나타내며, 단위 프레임의 에너지 세기를 직관적으로 정량화하는 특성이다. ZCR (Zero Crossing Rate)은 단위 프레임의 신호 시퀀스가 영교차점을 지나는 빈도수를 나타낸다.

본 연구에서는 추출된 음향 특성의 통계적 특성 또한 고려하기 위해, AE를 제외한 모든 음향 특성의 평균을 측정하여, 평균으로부터의 제곱거리 및 일반화 거리 또한 특성으로 활용하였다.

4. 분류 시스템 설계

본 절에서는 앞서 언급한 음향 특성을 바탕으로 분류 시스템의 설계를 보인다. 우선, 각 음악으로부터 추출되는 프레임의 개수가 길며 천차만별이기에, 이를 일반화하고 차원을 감쇄하기 위한 통계적 방법들을 고려하였다. 또한 차원 감쇄된 음향 특성을 다양한 분류기에 적용할 수 있는 스키마를 설계하였다. 설계한 스키마는 그림 2에서 보이고 있다.

4.1. 차원 감쇄 기법

기존의 음악 연구에서 사용한 프레임화 및 본 연구에서 사용하는 비트 기반의 프레임화 방법은 각 음악에서 매우 길며 일정치 않은 개수의 프레임을

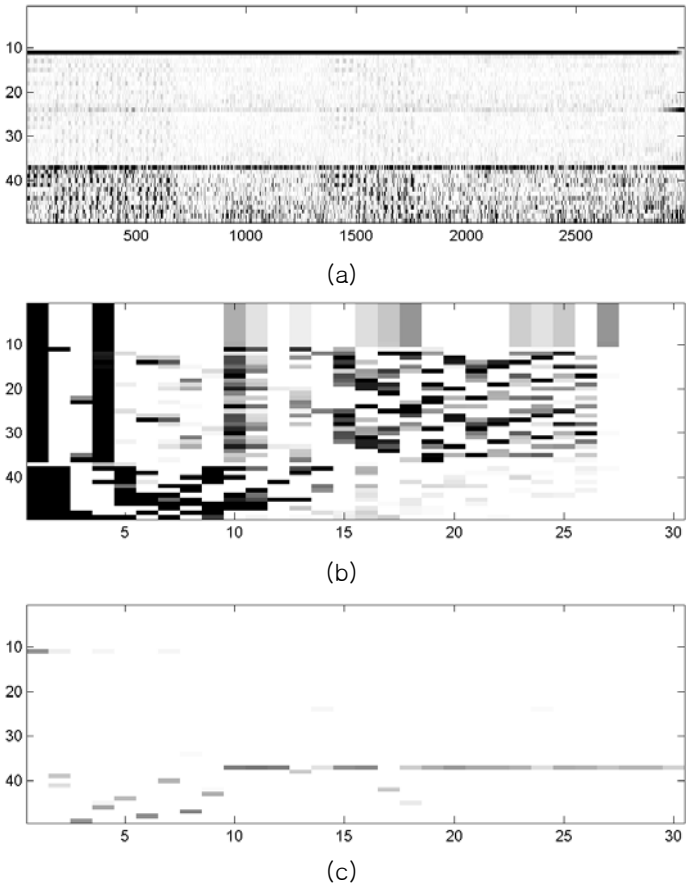

그림 3 (a) 차원 감쇄 전의 특성 벡터 (b) 시간 축을 30차원으로 감쇄한 PCA 결과 (c) 시간 축을 30차원으로 감쇄한 NMF 결과

추출한다는 단점을 가지고 있다. 가령, 140 BPM의 5분 길이 음악의 경우, 제안된 프레임화 방법에 의하면 5,600개의 프레임이 생성된다. 또한, 각 프레임에서 49개의 음향 특성이 추출되는 것을 감안하면, 해당 음악을 나타내기 위한 음향 특성 벡터는 약 27만 차원을 가지는 것을 알 수 있다. 따라서 다양한 프레임 개수를 표준화하고, 프레임 개수에 따라 거대해지는 음향 특성 벡터의 차원을 감쇄하기 위한 차원 감쇄 기법의 적용이 필요하다.

본 연구에서는 대표적인 차원 감쇄 기법인 PCA와 NMF를 적용하였다. 그림 3는 차원 감쇄 결과를 보이고 있다. PCA (Principal Component Analysis) [8]는 다양한 분야에서 활용되고 있는 대표적인 차원 감쇄 기법이다. PCA는 한 데이터를 표현하는 기반 벡터인 고유 벡터(eigenvector), 그리고 이를 복원하기 위한 계수로 분리할 수 있다. 그림 3 (b)은 PCA에 의한 차원 감쇄 결과를 보이고 있다.

한편 NMF (Non-negative Matrix Factorization) [9]는 부분 특징에 근거한 인식을 두뇌가 수행한다는 아이디어에 착안한 알고리즘이다. NMF는 회귀 훈련 과정을 통해 계산 가능하며, NMF 를 적용하기 위한 제약 조건은 원 데이터가 모두 음 아닌 값을 가져야 한다는 것이다. 따라서 본 연구에서는 NMF 적용 시 음향 특성 벡

터의 값들의 분포를 변환하여 모든 값들을 음 아닌 값으로 치환하였다. 그림 3 (c)은 NMF에 의한 차원 감쇄 결과를 보이고 있다.

4.2. 분류기 (classifier)

본 연구에서는 대표적으로 사용되는 분류기인 NN (신경망 네트워크), GMM (가우시안 혼합 모델), 그리고 SVM (Support Vector Machine)을 사용하여 분류를 진행하였다[10].

NN (신경망 네트워크)는 수많은 분류 연구에서 다양하게 활용되어 온 알고리즘으로, 생물학적 관찰에 의한 모델링과 직관적인 사용법이 강점이다. 그러나 입력 벡터 차원에 따른 은닉층 (hidden layer)의 적절한 디자인이 필요하다. 또한 BP (Back-Propagation) 알고리즘이 국부 최소 (local minimum)에 빠지지 않도록 초기값 및 훈련 과정의 제어가 필요하다.

GMM (가우시안 혼합 모델)은 주어진 데이터의 분포 밀도를 복수 개의 확률밀도함수로 모델링하는 방법이다. 이를 학습하기 위한 대표적인 방법으로 EM 알고리즘을 적용한다. GMM을 이용하여 확률밀도함수를 성공적으로 모델링 하기 위해서는 적절한 확률밀도함수 개수를 지정하는 것이 필요하다.

마지막으로 SVM은 커널 트릭 (kernel trick)을 이용하여 원 데이터를 가상의 고차원 공간으로 옮기고, 해당 공간에서 원 데이터를 최적으로 분류할 수 있는 최대 마진을 갖는 초평면(hyperplane)을 구하는 방법이다. 데이터에 따라 초고차원으로 차원이 확장될 수 있기에 수행 시간이 오래 걸릴 수 있다는 단점이 있지만, 좋은 성능으로 인해 최근 많은 연구에서 활용되고 있는 분류기이다.

5. 실험 결과

본 절에서는 그림 2의 스키마와 다양한 차원 감쇄 기법 및 분류기를 이용하여 훈련 및 검증한 결과를 보이고 해석한다.

5.1. 분류 시스템 구현

음향 특성 벡터 추출 시 일시적으로 30만 차원이 넘어가는 데이터가 추출될 수 있기에, 원활한 실험을 위해 실험은 모두 64-bit 환경에서 이루어졌다. OS로는 Windows Vista Enterprise K 64-bit edition를 구비하였으며, 실험 플랫폼으로는 MATLAB R2007b 64-bit을 활용하였다. 또한 펜티엄D 스미스필드 3.4GHz 듀얼 코어 환경에서 4GB 물리 메모리와 Western Digital Raptor 150GB 4개를 이용한 RAID 5 저장 환경을 적용하여 고차원의 특성 계산 시 빠른 메모리 스왑이 이루어지도록 구성하였다.

일반적인 대중 음악은 하나의 비트를 8~16 부분으로 쪼개어 표현하므로, 본 논문에서는 한 프레임을 1/8 비트 단위로 지정하였다. 또한, 일반적으로 주요한 고조파의 특성은 기본 고조파 및 배수 고조파를 합쳐 총 6개의 고조파에서 나타나므로, HDS에서는 고조파를 6개까지 통합하도록 지정하였다(M=6). MFCC의 경우 초기 13개의 벡터에서 주요 특성이 나타나므로, 에너지 특성이 두드러지게 나타나 AE와 중첩이 일어나는 제1벡터를 제외한 2~12번째 계수를 음향 특성으로 활용하였다.

표 2 STFT에 기반한 훈련 및 검증 결과 (%)

	NN	GMM	SVM
PCA	73.12	63.44	79.57
NMF	75.27	64.52	78.49

표 3 HDS에 기반한 훈련 및 검증 결과 (%)

	NN	GMM	SVM
PCA	73.12	64.52	79.57
NMF	74.19	64.52	80.65

음향 특성 벡터는 차원 감쇄에 의해 최종적으로 49 X 30 = 1,470 차원의 벡터를 가진다. NN이 최적의 해를 가질 수 있도록 은닉층의 노드 개수를 3천개로 지정하였다. 또한, GMM에 의해 디테일 있는 확률밀도함수가 형성되도록 혼합수를 30으로 지정하였다. 마지막으로 SVM은 LIBSVM v2.86 엔진[11]을 활용하였으며, SVM type으로 C-SVM, 그리고 커널 함수로 polynomial을 사용하였다.

5.2. 실험 데이터

분류 실험을 위한 데이터로 MIR 커뮤니티 및 MIREX 콘테스트, ISMIR과 같은 MIR 학회에서 다양하게 활용되고 있는 Allmusic.com[12]의 무드 분류 기준을 활용하였다. Allmusic.com은 약 190여 가지의 무드를 기준으로 음악을 분류하여 서비스하고 있으며, 이 중 대표적이며 무드 간의 경계가 극명하게 갈리는 분노 (angry), 지루한 (bored), 적막한 (calm), 흥분되는 (excited), 행복한 (happy), 신경질적인 (nervous), 평화로운 (peaceful), 기쁜 (pleased), 이완된 (relaxed), 슬픈 (sad), 졸린 (sleepy) 과 같은 11가지 무드 분류 기준에 맞는 93개의 음악을 수집하였다.

5.3. 정확도 측정 결과

표 2 및 표 3은 제안한 분류기에 의한 훈련 및 검증에 의한 정확도 실험 결과를 보이고 있다. STFT에 기반한 실험 및 HDS에 기반한 실험 모두 SVM이 가장 좋은 성능을 보였고, NN은 평균적인 성능을 보였으며, GMM은 좋지 않은 성능을 보였다. 그러나 이러한 성능의 차이는 분류기의 설정에 따라 달라질 수 있는 부분이다. 따라서 GMM이 SVM이나 NN보다 성능이 좋지 않다고 볼 수는 없다.

그러나 STFT 대신 HDS를 기반으로 음향 특성 벡터를 추출하였을 경우 정확도가 향상한 것을 볼 수

있다. 비록 NN의 경우 NMF에서 성능이 오히려 감소한 것으로 나오지만, GMM이나 SVM의 경우 전반적으로 성능이 향상된 것을 볼 수 있다. 이는, STFT 및 HDS 모두 음향 특성 벡터를 추출하기 위한 방법이 동일하였으므로, 기존의 STFT 대신 HDS를 사용하여도 성능 향상이 가능하다는 결론에 이를 수 있다.

마지막으로 NMF의 정확도 성능이 PCA보다 전반적으로 높음을 볼 수 있다. STFT 기반의 SVM 실험에서는 NMF의 정확도가 PCA의 정확도보다 낮지만, 다른 모든 실험에서는 NMF 기반의 차원 감쇄 기법이 PCA와 동일하거나 오히려 정확도를 좀더 높여주는 것으로 보이고 있다.

6. 결 론

본 논문에서는 새롭게 제안한 프레임화 기법과 기존의 스펙트럼 분석 방법인 STFT를 대체하기 위한 HDS를 음향 특성을 분석하기 위한 기본 토대로서 제안하였다. 또한 PCA, NMF를 비롯한 다양한 차원 감쇄 기법과 NN, GMM, SVM 등의 다양한 분류기를 사용하여 정확도 실험을 수행한 결과, HDS가 STFT보다 음향 특성 벡터 추출 시 좀더 정확도를 올릴 수 있으며, PCA보다는 NMF가 차원 감쇄 기법으로서 성능이 더 좋다는 것을 알 수 있었다.

향후 연구로서 새롭게 제안한 HDS를 기반으로 고조파 특성을 살리고, 이를 쉽게 추출하기 위한 연구가 필요할 것이다. 또한 기존의 다양한 프레임화 기법을 연구하고, 음악적 특성에 맞는 또다른 프레임화 기법을 사용하여, 차원 감쇄 시 음악적 정보를 잃지 않도록 하기 위한 연구가 필요할 것이다. 마지막으로, 보다 다양한 최신의 차원 감쇄 기법과 SVR, AdaBoost 등의 최신 분류기를 사용한 실험을 통해 음악 무드 분류를 위한 최적의 분류 조합을 찾아내는 연구가 필요할 것이다.

참고 문헌

[1] MIREX (Music Information Retrieval Evaluation eXchange), http://www.music-ir.org/mirexwiki/

[2] L. Lu, D. Liu, H. Zhang, “Automatic Mood Detection and Tracking of Music and Audio Signals,” IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.1, pp. 5—18, 2006.

[3] D. Ellis and G. Poliner, “Identifying `Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking,” in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), vol.4, pp. 1429—1432, 2007.

[4] William A. Sethares, “Tuning, Timbre, Spectrum, Scale,” Springer-Verlag, 2nd edition, 2005.

[5] Anssi Klapuri and Manuel Davy, “Signal Processing Methods for Music Transcription,” Springer-Verlag, 2006.

[6] T. Li and M. Ogihara, “Toward Intelligent Music Information Retrieval,” IEEE Transactions on Multimedia, vol.8, no.3, pp. 564—574, Jun. 2006.

[7] B. Logan, “Mel frequency cepstral coefficients for music modeling,” in Proc. of International Symposium of Music Information Retrieval (ISMIR), 2000.

[8] K. Pearson, “On Lines and Planes of Closest Fit to Systems of Points in Space,” in Philosophical Magazine, vol.2, no.6, pp. 559—572, 1901.

[9] D. D. Lee and H. S. Seung, "Learning the Parts of Objects By Non-negative Matrix Factorization," in Nature, vol. 401, pp. 788—791, 1999.

[10] Richard O. Duda, Peter E. Hart, David G. Stork, “Pattern Classification,” Wiley Press, 2nd edition, Oct. 2000.

[11] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines, 2001 Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[12] Allmusic.com, http://www.allmusic.com/