

음악의 대표구간을 이용한 내용기반 장르 판별에 관한 연구

The Content-based Genre Classification using Representative Part of Music

이종인 · 김병만
Jong In Lee and Byeong Man Kim

금오공과대학교 소프트웨어공학과

요 약

일부 음악 장르 분류에 관한 기존 연구에서는 특징 추출을 위한 구간 선택 시 사람이 직접 음악의 주요 구간을 지정하는 방법을 사용하였다. 이러한 방법은 분류 성능이 좋은 반면 수작업으로 인한 부담으로 새롭게 등록되는 음악들에 대해 지속적으로 적용하기가 곤란하다. 이러한 이유로 최근 음악 장르 분류와 관련된 연구에서는 자동으로 추출구간을 선정하는 방법을 사용하고 있는데 이러한 연구의 대부분이 고정된 구간 (예, 30초 이후의 30초 구간)에서 특징을 추출하는 관계로 분류의 정확도가 떨어지는 문제점을 갖고 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 음악 전체 구간에 대하여 반복구간을 파악하고, 그 중 음악을 대표할 수 있는 단일 대표구간을 선정한 후, 대표구간으로 부터 특징을 추출하여 장르 분류시스템에 적용하는 방법을 제안하였다. 실험 결과, 기존 고정구간을 사용한 방법에 비해 괄목할 만한 성능 향상을 얻을 수 있었다.

키워드 : 음악 장르 분류 시스템, 음악 대표 구간 선정, 음악 구조 분석, 음악 세그먼트 분류, 내용 기반 음악 특징 추출

1. 서론

최근 멀티미디어 데이터베이스에 저장된 음악 정보가 급속히 증가하고 있으며, 인터넷을 이용하여 디지털 음악을 수집하고 소장하는 사용자들이 증가하고 있다. 이에 따라 사용자들은 방대한 양의 디지털 음악을 효과적으로 분류 및 검색할 수 있는 시스템을 요구하고 있다. 그러나 지금까지의 음악 데이터들은 수작업을 통해 제목, 가수 등의 텍스트 기반의 인덱스가 입력되고 관리되어 오고 있다. 이러한 텍스트 기반의 검색 시스템은 사용자가 원하는 데이터를 찾기도 수월하지 않고, 메타데이터의 정확성에 의존하게 된다. 이에 최근에는 음악 내용을 기반으로 한 분류 및 검색 기술이 점차 중요시되어지고 있다.

음악 분류 중 일반적인 분류라 할 수 있는 음악 장르 분류에 관한 연구는 1990년대 후반, 2000년대 초반에 걸쳐 많은 연구들이 이루어져 오고 있다. 하지만 일반적으로 우리가 접하고 있는 음악 장르의 종류는 문화적, 음반 산업의 요구, 각각의 음악 그룹의 범위와 관계하여 다양한 분류법들이 존재하고 있고, 이러한 장르 분류법을 그대로 내용 기반의 분류 및 검색에 적용시키기에는 무리가 있다. 이에 대한 대안으로 내용기반 장르 판별, 즉 음악 오디오로부터 특징들을 추출하고 이를 기반으로 분류하는 연구들이 많이 이루어져 왔고, 그 결과, 음악 특징 추출에 대한 연구가 일정 수준 이상에 이르게 되었다.

음악은 전체적인 흐름에 있어 그 변화량이 다양하여 지금까지 나온 특징들을 사용하여 전역적인 특징을

추출하기 위해서는 연산량이 매우 많아지거나, 전역적인 특징을 담기 위한 저장소의 공간적 문제가 발생할 수 있고, 또한 하나의 장르에 대한 음악 내에서 다양한 변화의 부분 때문에 다른 장르의 특색 또한 포함되어 이를 분리/판별하는 것이 어렵다. 기본적으로 지금까지의 음악 장르 분류에 관한 연구들은 음악 내 일정 구간을 선택하여 그 구간 내에서 특징을 추출, 판정하는 절차를 거치게 되는데 그 구간을 결정함에 있어 전문가가 수작업으로 선택하거나 임의의(시작 후 30초 지난 지점부터 30초 구간) 구간을 선택하여 이루어지고 있다.

만약 음악내부의 구조 분석을 통하여 음악 내에서 이 음악을 대표 할 수 있는 구간, 즉 대표 구간을 추출하여 이를 이용한 내용 기반 장르 판별을 하게 된다면 전문가의 수작업 같은 많은 시간이 소모되는 작업이 생략되거나, 임의의 구간을 선택함에 따라 발생하는 오류에 대하여 보완하는 이점이 있으리라 본다. 이러한 이유에서 본 논문에서는 음악내의 전체적인 구조, 즉 반복되는 구간의 검출을 통한 음악의 반복구간 검출 후, 이 중 음악 전체를 대표할 구간을 선정, 이 대표 구간에서 특징을 추출하고 이를 자동 분류 실험에 사용하는 방법을 제안하였다.

2. 관련연구

2.1. 음악 장르 판정 연구들

음악 장르 판정에 관한 연구들은 대체로 장르 판정에 적합한 특징들에 대한 연구들이 대다수를 이루고 있

다. 그 중 [1]은 기존 음성인식 및 음향판별 분야에서 널리 쓰인 특징들인 음색(Timbre) 특징과 음악적 특징인 비트 히스토그램(Beat Histogram)과 피치 히스토그램(Pitch Histogram)을 사용하였다.

[2]에서는 음성인식 분야에서 시작된 MFCC보다 음악의 분류를 스펙트랄 대비(Spectral Contrast)라는 새로운 특징을 제안하여 5개의 장르에서 MFCC와 비교하여 더 좋은 성능을 나타내는 것을 보여주었다.

[3]에서는 DWCHs라 불리는 웨이블릿 히스토그램에 기초한 새로운 음악 특징 추출 방법을 제안하여 실험한 결과, [1]에 사용한 특징 중 음색 특징과 함께 DWCHs를 함께 사용 하였을 시, 장르 판정에 있어 두 개의 장르분류에서는 최대 99%의 성능을 보였고 10개의 장르분류에 적용하였을 때에도 78.5%의 높은 성능을 보여주었다.

Li와 그의 동료는 [4]에서 [3]의 연구를 바탕으로 음악 장르 판별을 위한 계층형 장르 분류법의 정리 필요성을 제기하였고, 기존 연구에서 사용한 장르 분류법을 2계층 장르 분류법에 적용하여 10개의 장르에 관해서 0.3%, 6개 장르 분류에 관하여 3%의 성능 향상이 있음을 보여주었다.

2.2. 음악 구조 분석을 통한 반복구간 탐지 및 추출

음악 내용 분석관련 연구들 중 정보검색분야의 연구와 함께 음악의 구조 분석에 관한 연구 또한 활발히 진행되어지고 있다. 이중 이미지의 썸네일(Thumbnail)처럼 음악에서도 적은 데이터를 가지고 음악에 대해 정보를 알 수 있는 부분의 추출에 관한 연구들이 진행되고 있다. [5]에서는 음악의 단일 썸네일이라 볼 수 있는 스니펫(Snippet)을 추출하기 위하여 크게 기본 특징 추출, 특이점 검출(Saliency Detection), 음악 구조 분석, 그리고 스니펫 생성이라는 네 가지 단계를 두어 스니펫을 추출하였다. 추출된 스니펫의 성능을 파악하기 위하여 개인적인 설문 평가로 좋음, 그럭저럭, 만족스럽지 못함을 나타내는 3, 2, 1의 세 가지 평가치를 사용하여 조사하였는데 전체적으로 2.62 정도의 만족할만한 평가를 얻었다.

[6]에서는 음악 특징을 고려하여 리듬과 하모니를 표현하기에 적합한 크로마 벡터(Chroma vector)를 기반으로 한 썸네일 추출 방법을 제안하였다. 이 연구에서는 음악의 절을 파악하기 위해 프레임 별로 크로마 벡터를 추출한 후, 크로마 벡터의 유사성과 연속성을 통하여 반복구간을 탐색하고 탐색된 반복구간들 간의 구조 분석 및 재구축을 통하여 다중 썸네일을 추출하였다.

3. 대표구간을 이용한 특징 추출

본 논문에서는 내용 기반 음악 장르 분류를 위하여 음악 전체 내용 중 대표 구간 하나를 선정하여 대표 구간 내의 특징들을 추출하고 이를 학습과 분류를 위한 특징으로 사용하는 방법을 사용한다. 전체 시스템 구조는 그림 1에서 보는 바와 같이 반복구간 탐색을 통한 음악의 구조 분석, 분석된 반복 구간 정보를 바탕으로 반복 구간 중 대표 구간으로 적합한 구간 선택, 대표 구간으로부터 음악의 특징 벡터 추출, 특징 벡터를 사용한 학습 및 판정의 절차를 거친다.

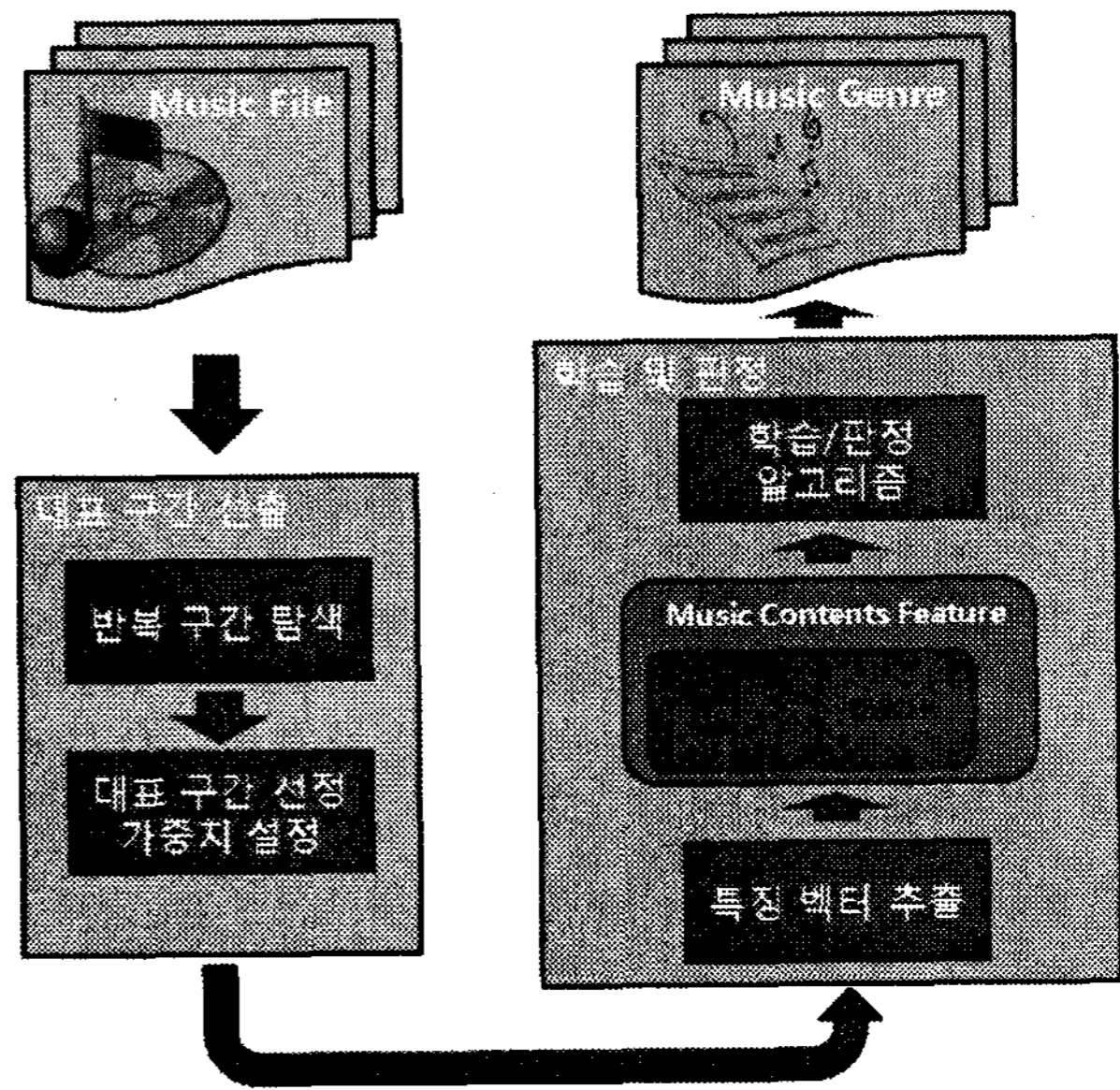


그림 1. 시스템 구조도

3.1. 반복 구간 탐색

음악의 대표 구간을 추출하기 위해선 우선 음악의 반복구간 탐색을 통한 구조 분석이 이루어져야 한다. 기본적으로 음악의 각 부분들은 소악절로 이루어져있고 이러한 소악절들 사이에는 반복적인 멜로디를 담고 있는 특징이 있다 [6]. 본 논문에서는 비슷한 멜로디로 이루어진 반복구간을 탐색하기 위해 [6]에서 사용한 12클래스의 피치(Pitch)분산 정보를 담고 있는 크로마 벡터를 사용한다.

◆ 12 피치 클래스의 크로마 벡터 계산

크로마 벡터는 각 프레임(250ms 크기) 별로 계산하게 되는데 각 프레임에서 추출된 파워스펙트럼($\Psi_p(f_{cent}, t)$)을 아래의 공식을 사용하여 12 피치 클래스(크로마)별 에너지를 구하게 된다 [7].

$$v_c(t) = \sum_{h=Oct_L}^{Oct_h} \int_{-INF}^{INF} BPF_{c,h}(f_{cent}) \Psi_p(f_{cent}, t) df \quad (1)$$

여기서, $BPF_{c,h}(f_{cent})$ 는 밴드 패스 필터로, 옥타브 h 에 해당하는 크로마 c 에 해당하는 주파수 영역을 추출하게 된다.

$$BPF_{c,h}(f_{cent}) = \frac{1}{2} \left(1 - \cos \left(\frac{2\pi(f_{cent} - (F_{c,h} - 100))}{200} \right) \right) \quad (2)$$

위 수식에서 사용된 $F_{c,h}$ 는 해당 크로마의 중심 값으로 센트 단위의 주파수 단위법을 사용할 경우 아래처럼 표현된다 (1옥타브 1200센트, 1크로마 100센트).

$$F_{c,h} = 1200h + 100(c + 1) f_{cent} \quad (3)$$

일반적으로 푸리에 변환 후 얻어진 스펙트럼의 주파수 단위는 헤르츠(Hertz) 단위로 표현되기 때문에, 위 기술된 공식에 적용하기 위하여 센트 단위로의 변환이

필요하다. 이를 위한 변환 공식은 아래와 같다.

$$f_{cent} = 1200 \times \log_2(f_{hz}/2^{\frac{3}{12}-5}) \quad (4)$$

◆ 크로마 벡터를 통한 반복 구간 연결

음악 전구간의 크로마 벡터를 추출한 뒤 반복 구간을 탐색하기 위해서 추출한 크로마 벡터를 5초 간격으로 3/4의 크기씩 중첩시키며 그룹화를 하게 된다. 그룹화 된 크로마 벡터 그룹을 사용하여 그룹간의 유사도를 연산한다. 그룹간의 유사도를 계산하기 위해 각 그룹간의 거리를 계산한다. 그룹간의 거리가 계산되어지면 시작 그룹(앞)에 가장 낮은 거리를 가지는 다음 그룹 4개와의 각 그룹간의 거리 T 를 기억하게 된다. 이는 음악 내 반복구간은 적어도 5번 이루어진다는 가정을 바탕으로 한다 [6].

그룹간의 T 의 값 들이 계산되면, 이를 이용하여 그룹 간 연결을 통한 반복구간을 연결 하게 된다. 이때 연결되는 그룹의 길이는 적어도 12.5초(4 그룹) 이상이 되게 한다.

◆ 대표 구간 선택

반복구간 탐색이 완료 된 후, 장르의 내용을 추출하기 위해 하나의 구간, 즉 대표 구간을 추출하여 사용하게 된다. 이를 위해 탐색된 반복 구간 중 대표구간으로 선택하기 위한 가중치를 설정하게 되는데, 대표 구간 선택을 위한 가중치는 구간의 위치와 에너지에 따른 가중치 $W = W_p \times W_E$ 를 적용시키게 된다.

기존의 음악 Thumbnail 추출관련 연구들에 따르면 대다수의 음악들의 경우 음악의 주요한 내용이 담기는 구간은 음악 전체 중 중앙 부분에 위치하는 경향이 있는 것으로 나타나고 있다 [5, 6, 8]. 이를 바탕으로 위치에 대한 가중치는 음악 전체 길이 S_t 의 중앙 위치인 $\frac{S_t}{2}$ 에 구간 위치 가중치(W_p)를 1로 하여 외부로 갈수록 점차 낮아져 시작부분과 끝부분은 0이 되는 Hamming윈도우를 사용하여 가중치(W_p)를 주게 된다. 또한 위치와 함께 구간의 평균 에너지를 가중치 W_E 로 사용하게 되는데 반복 구간 중 낮은 에너지를 가지는 부분은 서주나 간주 부분이 검출되어있을 수 있고, 음악 내에 하이라이트 구간이 다른 구간에 비해 에너지가 높다는 가정 하에 적용시키는 것이다.

3.2. 특징 벡터 추출

대표 구간이 추출되어졌다면 내용 기반 장르 판별을 위해 음악의 특징 벡터를 추출해야 한다. 본 연구에서는 기존 연구들에서 사용된 음악의 특징들을 특징 벡터의 요소로 선택하였다. 각 특징들은 기존 연구들 및 일반적으로 널리 알려져 있으므로 간략한 특징에 대해서만 서술하였다.

◆ 음색 특징 (Timbral Texture Features)

음색 특징을 나타내는 특징으로는 크게 주파수의 외형(Spectrum Shape)를 나타내는 스펙트럼의 분산학적 특징, 사람의 청각 모델에 기반 한 MFCC 그리고 옥타브 밴드 기반의 음색의 대비를 표현하는 Spectral

Contrast 세 종류로 나누어 사용하였다 (표 1 참고).

표 1. 한 프레임에서 추출되는 음색 특징의 종류와 수

특징군	특징	특징수
Spectral Shape	Centroid	1
	Spread	1
	Roll-off	1
	Flux	1
Spectral Contrast	Valley	6
	Contrast	6
MFCC	Order 0~12	13
Total		29

▶ Texture Window.

일반적인 STFT(Short-Time FT)를 사용하여 음색 특징을 추출할 경우 음향 신호의 변화량이 많은 특징에 따라 프레임 간 특징 계수의 분산의 차이가 심하게 나타나 표준 분산으로 모델링하기에 무리가 있다. 이를 위해 [1]에서는 Texture Window와 Analysis Window 두 개의 윈도우를 사용하여 음색 특징을 추출하였다. 본 연구에서도 이 Texture Window를 적용시켜 음색 특징을 추출하였다. 음색 특징추출을 위한 파라메타와 방법은 다음과 같다.

- i) 23ms의 Analysis Window를 20ms씩 이동시키며 각 프레임별 음색을 추출
- ii) 0.6s의 Texture Window를 0.5s씩 이동시키며 Texture Window내의 Analysis 프레임을 그룹화. 각 그룹별 평균과 표준 편차를 계산
- iii) 추출 구간 내의 Texture Window 그룹의 평균과 표준편차계산 (총 특징 차수 29 * 4)

◆ DWCHs

DWCHs는 [3]에서 제안된 음악의 특성으로 옥타브 기반의 주파수 대역과 상응하는 다 해상도 웨이블릿 변환의 특성을 음악의 특징으로 사용하기 위하여 음원을 다우비치 웨이블릿 필터(Db8)를 사용하여 7레벨 변환 후 각 밴드별 웨이블릿 계수들을 히스토그램 화하여 각 기 첫 3개의 모멘트와 밴드별 에너지를 추출하여 사용하였다 (특징 차수 28).

4. 음악 장르 분류 실험

4.1. 실험의 구성

분류된 장르의 음악 데이터베이스를 구축하기 위하여 2005 MIREX [9]에서 제공한 음악 데이터들과 인터넷으로부터 수집한 MP3파일을 AllMusic [10]의 분류에 맞춰 8개의 장르로 분류하였다. 8개의 장르는 클래식(Classic), 일렉트로닉(Electronic), 재즈(Jazz), 메탈(Metal), 리듬 앤 블루스(R&B), 소프트 락/팝(Rock) 그리고 월드(World Music) 이며, 각 장르별 약 100개의 파일로 구성 총 700개의 음악파일을 사용 하였다. 데이터베이스 내의 모든 MP3파일의 샘플링레이트는 44100Hz로 고정하고 특징을 추출하기 위한 전처리 과정으로 모든 음악의 채널을 모노 채널로 변환하여 사용

하였다.

내용 기반 분류를 위한 학습 알고리즘으로 일반적으로 학습에 있어 성능이 좋기로 알려진 SVM, 다중 신경망(NN) 그리고 k-NN 알고리즘을 사용하였고, 각 학습 알고리즘을 적용시킨 성능을 평가하기 n-Fold Cross-Validate(5-fold)를 사용하여 성능을 평가하였다. 각 알고리즘에 대한 실험은 Weka라이브러리 [11]를 사용하였다.

4.2. 반복구간을 이용한 대표구간 선정과 대표구간을 이용한 장르 판정 성능

음악 전체를 대표하는 구간을 선정하기 위해 3.1 절의 방법을 통해 대표 구간을 선정하고, 대표구간으로부터 음악의 특징 점들을 추출하여 장르 판정에 사용하였다. 성능 비교를 위해 음악의 고정된 구간인 전체 길이의 20%, 50%, 80% 지점과 탐색되어진 반복 구간 중 대표구간으로 선정된 지점이 아닌 가장 처음 반복구간과 마지막 반복구간으로부터 특징 점들을 추출할 경우의 비교 실험을 하였다. 비교 결과는 [표 2]와 같다.

표 2. 추출 구간에 따른 성능 비교

		SVM	NN	k-NN(1)	k-NN(5)
고정 구간	20%	66.0969	66.0969	53.2764	56.9801
	50%	66.3818	65.3846	57.9772	60.6838
	80%	66.9516	68.6610	55.6980	58.8319
반복 구간	First	57.1633	59.0258	44.5559	47.2779
	Last	66.7622	65.3295	52.7221	57.3066
	Select	69.6275	72.0630	64.3266	64.8997

[표 2]에서 보는 바와 같이, 본 연구에서 제안한 방법인 대표구간 선택(Select)에 의한 장르의 분류가 고정위치에 비해 좋은 성능을 보임을 알 수 있다, 고정위치 중 최대 성능을 보이는 80% 구간의 특징을 사용한 다중신경망 성능 68.6610%보다 약 4% 향상된 72.0630% 분류 성능을 보여주고, 다른 학습 알고리즘 적용 시, 또한 3~7% 정도의 높은 분류 성능을 보여주고 있다. 특히 고정 구간을 사용하였을 때 다른 학습 알고리즘에 비해 성능이 비교적 낮았던 k-NN알고리즘의 경우에도 대표 구간을 사용할시 성능의 향상이 많았다. (k-NN(1)의 경우 고정구간 중 가장 높은 성능을 나타내는 50%의 지점에 비해 약 7%의 성능 향상.).

반복구간 중 선택 구간에 대한 실험에서는 반복구간의 첫 구간을 사용한 경우가 가장 성능이 낮게 나왔다. 이는 반복구간 탐색 시, 비슷한 멜로디를 지니고 악기의 연주로만 이루어진 서주(Intro) 부분과 간주 부분이 반복구간으로 탐색되고, 곡의 앞부분에 위치한 서주구간이 첫 번째 반복구간으로 선택되기 때문이다. 일반적으로 서주 구간은 장르별 판별능력이 낮은 결과를 보여준다 [8].

5. 결론

본 논문에서는 내용기반 장르 분류를 위하여 음악 내 반복구간 검출을 통한 대표 구간 선택에 대해서 살펴보았다. 음악의 반복구간 검출을 위한 연구의 경우

음악의 썸네일 추출이나 지능형 구간 탐색과 같은 음악 청취자나 검색자를 위해 연구가 진행되어 왔으나 이를 사용한 지능형 검색 분야에 적용한 사례는 없어왔다. 이를 장르 판정시스템에 적용시켜 사용한 결과, 음악의 고정된 구간에서 추출한 결과와 비교하였을 때 높은 성능, 특히 국지적 밀도와 관계한 k-NN 알고리즘에 있어 더 좋은 분류 성능을 보여주었다.

향후, 본 연구의 결과, 즉 반복구간을 통한 대표구간 선정 시 내용 기반 음악 분석 및 분류에 있어 더 나은 결과를 유도할 수 있다는 결과를 바탕으로 단일의 대표구간이 아닌 전역적인 구간 및 구조를 적용한 내용기반 음악 검색 및 추천 시스템에 관한 연구를 진행하고자 한다.

참고문헌

- [1] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", IEEE Transactions on Speech and Audio Processing, vol. 10, Issue 5, pp. 293-302, July 2002.
- [2] D. Jiang, L. Lu, H. Zhang, J. Tao, L. Cai, "Music type classification by spectral contrast feature", in Proc. of ICME '02, vol.1, pp. 113-116 2002.
- [3] T. Li, M. Ogihara, Q. Li, "A comparative study on content-based music genre classification", in Proc. of the 26th ACM SIGIR conference, pp. 282-289, 2003.
- [4] T. Li, M. Ogihara, "Music genre classification with taxonomy", in Proc. of ICASSP '05, vol. 5 pp. 197-200, March 2005.
- [5] L. Lu, H. Zhang, "Automated extraction of music snippets", in Proc. of the 11'th ACM international conference on Multimedia, pp. 140-147, 2003.
- [6] T. Zhang, R. Samadani, "Automatic Generation of Music Thumbnails", in Proc. of IEEE International Conference on Multimedia and Expo 2007, pp. 228-231, July 2007.
- [7] M. Goto, "SmartMusicKIOSK: music listening station with chorus-search function", in Proc. of the 16th UIST, pp. 31-40, November 2003.
- [8] 윤원중, 이강규, 박규식, "Multi-Feature Clustering을 이용한 강인한 내용 기반 음악 장르 분류 시스템에 관한 연구", 전자공학회 논문지, 제 42권, 3호, pp. 115-120, 2005. 05.
- [9] <http://www.music-ir.org/mirex/2005>.
- [10] <http://www.allmusic.com/>
- [11] <http://www.cs.waikato.ac.nz/ml/weka/>