

논문 2006-43SP-1-10

# 내용기반 음악 검색 시스템에서의 검색 속도 향상에 관한 연구

( A Study on the Retrieval Speed Improvement from Content-Based  
Music Information Retrieval System )

윤 원 중\*, 박 규 식\*\*

( Won-Jung Yoon, and Kyu-Sik Park )

## 요 약

본 논문에서는 빠르고 안정적이면서도 높은 검색 성공률을 보장하는 내용기반 음악 정보 검색 시스템을 구축하였다. 시스템 질의 구간이나 질의 길이에 따른 시스템 불안정성 문제를 해결할 수 있는 DB 구축 방법인 MFC기법과 각 SuperClass별로 특징 벡터의 차수를 차등 적용하여 시스템의 검색 속도를 향상시킬 수 있는 기법을 적용하였다. SuperClass를 적용한 시스템은 SuperClass를 적용하지 않은 시스템과의 검색 성공률, 검색 속도 그리고 검색 Precision 비교 실험에서 대등한 성능을 유지하면서 검색 속도를 20%~40% 향상시켰다.

## Abstract

In this paper, we propose the content-based music information retrieval system with improved retrieval speed and stable performance while maintaining reasonable retrieval accuracy. In order to solve the in-stable system problem, multi-feature clustering (MFC) is used to setup robust music DB. In addition, the music retrieval speed was improved by using the SuperClass concept. Effectiveness of the system with SuperClass and without SuperClass is compared in terms of retrieval speed, accuracy and retrieval precision. It is demonstrated that the use of MFC and SuperClass substantially improves music retrieval speed up to 20%~40% while maintaining almost equal retrieval accuracy.

**Keywords :** Music Information Retrieval, Multi-Feature Clustering(MFC), SuperClass, Retrieval Speed Improvement.

## I. 서 론

내용 기반 음악 장르 분류 및 검색은 지난 몇 년간 꾸준히 관심을 모으고 있는 연구 분야로써 음악 디지털 라이브러리(MDL), 방송·연예 산업, 가상현실 그리고 웹 어플리케이션 등 활용분야가 무궁무진하다 할 수 있

다. 이러한 내용 기반 음악 정보 검색 시스템에 대한 연구는 국내·외적으로 활발히 진행되어 이미 다양하면서도 상당한 수준의 연구 결과들이 쏟아지고 있다.

Tzanetakis와 Perry는 [1]에서 음악 신호로부터 음색 특징, Wavelet Transform기반의 리듬 정보 그리고 피치 정보에 해당하는 특징 벡터를 추출 조합하여 약 74%의 성공률을 보였다. Guo 와 Li[2] 는 SVM (Support Vector Machine) 기반의 이진트리 기법을 이용하여 검색 오류율을 줄일 수 있는 방법을 제안하였다. Subramanya는 [3]에서 각 장르별 Centroid를 구하여 장르를 대표하는 표준 벡터를 생성하고, 표준 벡터와 질의 음악의 Euclidean Distance(ED)를 측정하여 질

\* 학생회원, \*\* 정회원, 단국대학교 컴퓨터과학 및 통계학과 (Dept. of Computer Science and Statistics, Dankook University)

※ 본 연구는 한국과학재단 목적기초연구(R01-2004-000-10122-0)지원으로 수행되었음.

접수일자 : 2005년7월12일, 수정완료일 : 2005년11월10일

의 음악의 장르를 결정하였다. 이 결과를 이용하여 각 장르를 몇 개의 sub-class로 구분하고 각 sub-class의 특징 벡터의 차수를 차등 적용하는 방법을 사용하여 검색 속도를 향상시켰다. 그러나, 사용하는 특징 벡터의 차수가 높고, 질의 음악의 길이가 20초 정도로 상당히 길어 300곡의 DB를 검색하는데 약 40초의 시간이 걸려 실제 시스템 구현에는 문제가 있다. 그 외 오디오 신호의 내용기반 장르 분류와 검색에 관련된 다양한 논문 [4]-[10]이 연구되어 왔다.

본 논문은 논문[10]에서 제안된 MFC(Multi-Feature Clustering)기법과 기존 Subramanya의 논문[3]에서의 시스템의 성능과 검색속도를 향상 시킬 수 있는 방법을 적용한 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 II장에서는 본 논문에서 사용된 특징들에 대한 간단한 설명을 하였으며, III장에서는 DB구축 방법과 검색 속도 향상 알고리즘에 대하여 설명하였다. IV장에서는 다양한 컴퓨터 모의실험을 통해 비교 분석을 하였으며, 마지막으로 결론으로 끝을 맺는다.

## II. 음악 특징 벡터 추출

각 특징 벡터의 분석 및 추출은 다음과 같은 조건에서 수행하였다. 먼저 음악 신호는 22050Hz, 16bits, mono로 샘플링 되었으며 실험에 사용된 음악 클립은 20초 분량에 해당한다. 다음으로는 각 음악 신호에 23ms 크기의 Hamming window를 인접한 프레임과 25%씩 중복되게 이동하면서 각 23ms 프레임으로부터 음악 신호의 음색 특징들과 계수 영역 특징 등 2가지 형태의 특징들을 추출하였다. 음색 특징들로는 spectral centroid, spectral rolloff, spectral flux 그리고 zero crossing rates를 추출하였으며 계수 영역 특징들로는 13차의 Mel-frequency cepstral coefficients(MFCC)와 10차의 linear predictive coefficients(LPC)를 사용하였다. 음악 파일의 각 프레임에서 이상의 6가지 특징들을 추출하여 평균과 분산을 구하면 총 54차의 특징 벡터가 구해진다. 각 특징 벡터를 구한 후 전체 특징 벡터를 평균이 0, 분산이 1이 되도록 정규화시켜 사용하였는데, 이는 각 특징 값들의 편차로 인한 실제 음악 장르 분류의 오동작을 방지하기 위함이다. 본 논문에서 사용된 특징 벡터들은 일반적인 신호 처리 분야에서 잘 알려진 것들이므로 상세한 설명은 배제하고 다음의 표 1에 간략한 정의만 소개하였다.

표 1. 음악 특징에 대한 정의  
Table 1. Musical feature definitions.

특징 계수	설 명
Spectral Centroid	STFT의 magnitude 스펙트럼의 중심을 뜻하며, 스펙트럼의 형태를 측정하는 방법 중의 하나이다.
Spectral Rolloff	스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지를 보여준다.
Spectral Flux	연속된 스펙트럼 분포에서 정규화된 magnitude 들 간의 차이를 제공해서 구할 수 있으며, 스펙트럼 변화의 양을 계산할 수 있다.
ZCR	오디오 신호 파형의 위상이 중심축을 통과하는 회수를 나타내며, 신호의 주파수 내용을 측정하는 가장 간단한 특징으로 음성인식에서 유·무성음의 판별에 사용된다.
MFCC	인간의 청각 특성을 모델링 하는 방법으로 오디오 신호의 magnitude 스펙트럼을 log scale 한 후 FFT bin을 그룹화하여 인간의 청각 특성에 맞는 Mel-Frequency scale로 변환한 것이다.
LPC	인간의 발성 모델에 입각해서 음성 신호를 부호화하는 방법으로 오디오 파형의 샘플 값에서 필터 계수를 구하여 성대에서 입, 코까지의 성도 특성을 8~12차의 전극형(All-pole) 필터에 근사시키는 방법이다.

## III. 제안된 알고리즘

### 1. Audio DB 구축 방법

본 논문에서는 음악 검색 시 전장에서 추출한 54차 특징들을 모두 사용하지 않는다. 전장에서 구한 특징들은 상호간의 상관도가 높고 또한, 시스템의 연산량이 높아짐으로, 효율적인 시스템을 구축하기 위해서는 전체 54차의 특징 벡터로부터 효과적인 특징들만을 선별해주는 작업이 필요하다.

본 연구에서는 본 연구팀이 [10]에서 제안한 sequential forward selection(SFS) 기법과 k-means clustering 알고리즘에 기반한 multi-feature clustering (MFC)기법을 사용하여 10차의 특징 벡터를 이용하여 DB를 구축하였다. MFC 알고리즘의 기본 동작은 다음과 같다. 먼저 DB로 구축될 각 음악 신호 전체 길이에 대해 20초 단위의 큰 윈도우를 쉼표가면서 각 윈도우에 SFS를 적용하여 최적의 특징 벡터 열을 추출한다. 다음으로는 이 결과로 얻어진 특징 벡터 열들을 k-means clustering 기법을 이용하여 최종 4개의 특징 벡터 열로 분류한다. MFC기법을 이용하여 DB를 구축하는 이유는 시스템 질의 구간이나 질의 길이에 따른 시스템 불안정성 문제를 해결할 수 있기 때문이다. 다음의 그림 1은 MFC 알고리즘의 전체 개요를 나타낸다.

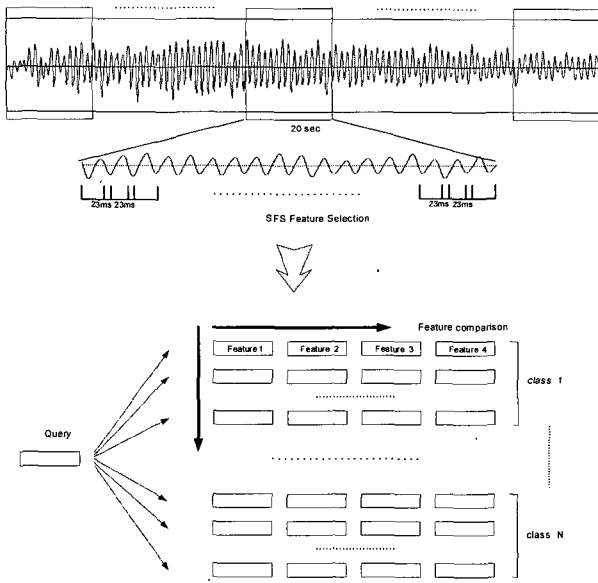


그림 1. 학습 DB에 대한 multi-feature clustering 절차  
 Fig. 1. Multi-feature clustering procedure for training database.

2. 장르 분류 및 검색

본 논문에서는 내용 기반 음악 장르 분류를 통한 음악 검색 시스템을 제안한다. 그러나, 단순히 장르 분류 결과를 이용하여 Query가 분류된 장르에서만 검색을 한다는 것은 상당한 리스크를 감수해야 한다. 만약, Query에 대한 장르 분류가 잘 못 되었다면, 시스템은 전혀 상관이 없는 장르의 DB만을 검색하여 쓸모없는 검색 결과를 사용자들에게 제공하게 될 것이다. 본 논문에서는 이러한 문제점을 해결하고자 다음의 방법을 사용한다.

우선 장르 분류를 위하여 DB의 각 장르별로 표준 벡터를 생성하며, Query는 각 장르의 표준 벡터와의 Euclidean Distance(ED)를 측정하여 확률로 변환한다. 이렇게 변환된 확률을 이용하여 해당 확률이 상위 30% 내에 있는 장르들을 SuperClass1, 다음 30%의 장르들을 SuperClass2 그리고, 나머지 확률의 장르들을 SuperClass3으로 우선순위를 결정짓는다. 다음으로 결정된 우선순위에 대하여 검색에 사용될 특징 벡터의 차수를 결정한다. 상대적으로 검색될 확률이 높은 SuperClass1에는 DB에 구축되어 있는 것과 동일한 차수의 특징 벡터를 사용하고, 검색될 확률이 조금 낮다고 판단된 SuperClass2에는 DB의 특징 벡터의 차수의 1/2을 사용하고, 마지막으로 검색될 확률이 가장 낮다고 판단된 SuperClass3에는 DB의 특징 벡터의 차수의 1/4을 사용하게 된다.

SuperClass를 사용한 경우와 SuperClass를 사용하지

않은 경우에 대한 검색 시간을 비교해보면, 한 곡당 모든 특징 벡터를 사용하여 구한 ED 계산 시간을  $\beta$ , DB의 곡 수를 N, 장르의 수를 M이라고 하면, 장르 분류를 하지 않고 모든 특징 벡터를 사용하여 검색을 하였을 경우  $\beta \cdot N$ 의 시간이 걸리게 된다.

SuperClass를 사용한 경우에는 시스템이 구축되어 이미 각 장르별 표준 벡터가 구해진 상태이므로 각 장르의 표준 벡터와의 ED를 구하는 시간  $\beta \cdot M$ 이 걸리며, SuperClass1에 대하여  $\beta \cdot 0.3N$ , SuperClass2에 대하여  $0.5\beta \cdot 0.3N$  그리고, SuperClass3에 대하여  $0.25\beta \cdot 0.4N$ 의 시간이 걸려  $(\beta \cdot M) + (\beta \cdot 0.3N) + (0.5\beta \cdot 0.3N) + (0.25\beta \cdot 0.4N) = \beta(M + 0.55N)$ 의 시간이 걸린다. 일반적인 시스템에서  $N \gg M$ 이므로, SuperClass를 사용할 경우 검색 시간을 약 40%정도 단축시킬 수 있다. 각 SuperClass별 특징 벡터의 차수 결정과 실험 결과는 IV장의 모의 실험에서 자세히 다루기로 한다.

표 2. SuperClass별 특징 차수 할당

Table 2. Feature vector allocation for the each SuperClass.

장르 추정 확률	구분	검색에 사용될 특징의 차수
70% 이상의 장르들	SuperClass1	DB와 동일한 차수 (10차 특징 벡터)
40% 이상의 장르들	SuperClass2	DB의 차수의 1/2
40% 이하의 장르들	SuperClass3	DB의 차수의 1/4

IV. 음악 검색 실험

1. 실험 구성

제안된 알고리즘의 실험을 위해 사용된 음악 DB는 인터넷 전문 음악 사이트의 MP3 파일, 음악 CD 그리고 라디오로부터 Classic, Hiphop, Jazz 그리고 Rock의 4개 음악 장르에 대해 각 장르별로 300곡을 선정하여 총 1200곡을 사용하였다. 실험에 사용된 음악 파일은 22050Hz, 16bits, mono의 wave 파일로 변환하여 사용하였다.

1200곡의 음악 파일들은 각 장르별로 무작위로 270곡씩 총 1080개(90%)의 학습 데이터를 수집하여 표준 벡터를 추출하였으며, 추출된 표준 벡터와의 거리를 기반으로 SuperClass들을 구성하였다. 검색을 위해서 음악 질의는 전체 DB 내의 음악 파일의 4개의 특징 벡터 세트들과 단순 Euclidean 거리를 측정하여 비교하며, 검색 성공률은 검색 결과 상위 10곡 안에 질의 음악이 포함되어 있다면 검색이 성공한 것으로 간주하였다. 또한,

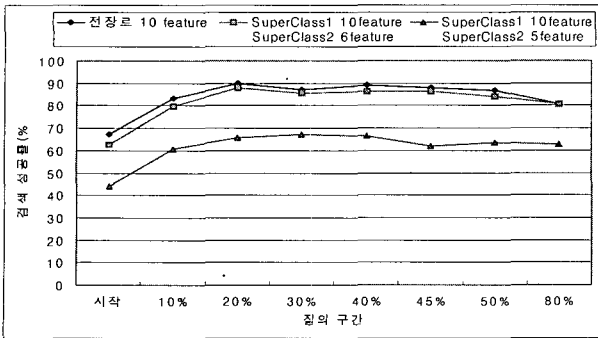


그림 2. SuperClass를 적용한 시스템과 SuperClass를 적용하지 않은 시스템에 대한 검색 성공률의 비교  
Fig. 2. Compare the retrieval accuracy for the system with SuperClass and without SuperClass.

제한된 DB에서 신뢰성 있는 결과를 도출하기 위해서 100번 반복하여 실험 결과를 도출하였으며 최종 성공률은 각 반복 실험의 성공률의 평균을 통해 얻었다.

다음 그림 2는 각 SuperClass의 적절한 특징 벡터의 차수를 찾기 위한 실험의 결과이다. 본 논문에서는 4개 음악 장르에 대하여 실험을 하였기 때문에 SuperClass1, SuperClass2로 구분을 하였다. 같은 음악 파일에서 5초의 절의를 총 8개의 다른 구간 - 음악 파일의 시작부분, 음악 파일 전체 길이의 10%, 20%, 30%, 40%, 45%, 50% 그리고 80% 지점 - 에서 추출하였다.

실험 결과 SuperClass를 적용하여 SuperClass2에서 특징 벡터의 차수를 6차까지 낮췄을 경우의 검색 성공률은 SuperClass를 적용하지 않고 모든 장르에서 10차의 특징 벡터를 사용하여 검색한 성공률과 차이가 없었으나, SuperClass2의 특징 벡터의 차수를 5차로 낮출 경우는 검색 성공률이 현저히 낮아졌다. 처음의 가정은 우선순위를 낮추면서 특징 벡터의 차수를 1/2로 줄이는 것이었으나, 시스템의 성능이 현저히 약화되어 SuperClass2에서의 특징의 차수를 60%로 낮추어 사용하기로 결정하였다.

정보 검색 시스템의 검색 결과는 다음 표 3과 같이 분류할 수 있다. 이렇게 분류된 정보들을 이용하여 시

표 3. 정보 검색 결과에 대한 분할표  
Table 3. Contingency table for the information retrieval result.

	관련 있는 것 (Relevant)	관련 없는 것 (Not Relevant)
검색된 것 (Retrieved)	a	b
검색되지 않은 것 (Not Retrieved)	c	d

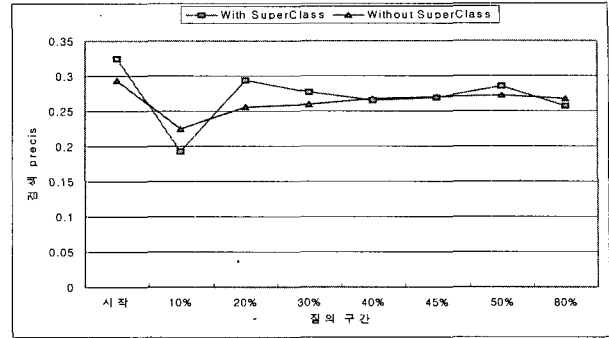


그림 3. 각 질의 구간에서의 precision 결과  
Fig. 3. Precision result from each query portions.

스템의 성능을 평가하는 척도로 Precision을 들 수 있다. Precision은 단순히 질의에 부합하는 유용한 정보를 찾았는지·못 찾았는지를 표시하는 검색 성공률(Retrieval Accuracy)과는 달리 검색된 결과들 중에 질의와 관련된 것들이 얼마나 있는가를 알려준다.

본 논문에서 구현한 시스템은 음악의 곡명과 장르만을 이용한 내용기반 음악 정보 검색 시스템으로 Precision은 검색 결과 중의 동일 장르 음악의 비율을 나타낸다고 할 수 있다.

다음 그림 3은 시스템의 Retrieval Precision 결과를 나타내고 있다. Precision을 통한 비교 실험에서도 SuperClass를 적용한 시스템의 성능이 SuperClass를 적용하지 않은 시스템의 성능과 큰 차이가 없음을 확인할 수 있다.

Precision은 식(1)과 같이 계산할 수 있다.

$$Precision = \frac{a}{a+b} \tag{1}$$

다음 그림 4는 본 논문에서 사용한 방법이 실제 시스템에서 검색 시간을 얼마나 단축시켜 주는지에 대한 실험 결과이다. DB를 240곡에서 1200곡까지 단계별로 늘

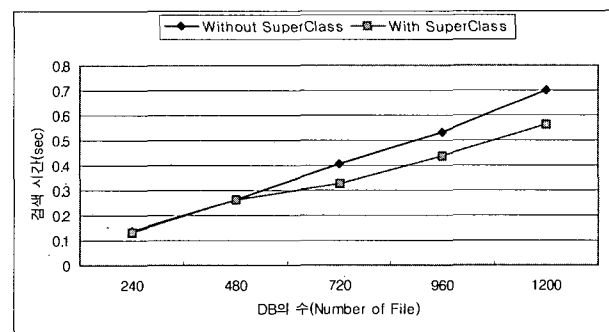


그림 4. DB의 증가에 따른 검색 시간  
Fig. 4. Compare the retrieval time for the increasing of DB size

려가면서 측정해보았다. DB size가 작을 경우에는 SuperClass를 적용한 시스템이나 SuperClass를 적용하지 않은 시스템이나 검색 시간의 큰 차이가 없으나 DB size가 커질수록 선형적으로 검색 시간이 단축되는 결과를 볼 수 있다. 그림 4에서는 2개의 SuperClass로 나누어 실험을 하여, 약 20%의 검색 시간이 단축되었음을 확인할 수 있다.

## V. 결 론

본 논문에서는 기존의 MFC 기법으로 구축된 DB를 이용하여 높은 검색 성능을 유지하면서도 검색 시간을 단축할 수 있는 SuperClass개념을 이용한 내용기반 음악 검색 시스템을 구축하였다. 본 논문에서는 5초 길이의 음악을 질의로 간단한 장르 분류를 한 후, 장르 분류 결과를 이용하여 상위 30%의 장르는 SuperClass1으로, 그 다음 상위 30%의 장르는 SuperClass2 그리고, 마지막 40%의 장르는 SuperClass3로 분류하여 해당 SuperClass마다 특징 벡터의 차수를 차등 적용하여 검색 시간을 단축하였다. SuperClass를 사용하지 않은 시스템에 비하여 약 20%~40%의 검색 시간을 단축할 수 있었다. 향후 본 연구 결과를 이용하여 잡음 환경의 모바일 기기에 적용 가능한 내용기반 음악 검색 시스템 구현을 위한 연구를 진행할 예정이다.

## 참 고 문 헌

- [1] G. Tzanetakis, "Manipulation, Analysis and Retrieval Systems for Audio Signals", Ph. D. Thesis in Computer Science from Princeton University, June, 2002.
- [2] P. G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machine", IEEE Trans. on neural networks, vol. 14, no. 1, pp. 209-215, Jan., 2003.
- [3] S. R. Subramanya, A. Youssef, B. Narahari, and R. Simha, "Automated Classification of Audio Data and Retrieval Based on Audio Classes", International Conference on Computers and Their Applications, Cancun, Mexico, April, 1999.
- [4] J. Foote et al, "An overview of audio information retrieval", ACM-Springer Multimedia Systems, vol. 7, no. 1, pp. 2-11, Jan. 1999.
- [5] S. Z. Li, "Content-based classification and retrieval audio using the nearest feature line method", IEEE Trans. on Speech Audio Processing, vol. 8, pp. 619-625, Sept., 2000.
- [6] Y. Wang, Z. Liu and J. Huang, "Multimedia content analysis: using both audio and visual clues", IEEE Signal Proc. Mag., Nov., 2000.
- [7] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machine", IEEE Trans. on neural networks, vol. 14, no. 1, pp. 209-215, Jan., 2003.
- [8] T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification", in Proc. of the 26th annual internal ACM SIGIR, pp. 282-289, ACM Press, July, 2003.
- [9] J. J. Burred and A. Lerch, "A hierarchical approach to automatic musical genre classification", in Proc. DAFx03, pp. 308-311, Sept., 2003.
- [10] Kyu-Sik Park, Won-Jung Yoon, Kang-Kue Lee, Sang-Heon Oh and Ki-Man Kim, "MRTB framework: a robust content-based music retrieval and browsing", Consumer Electronics, IEEE Transactions on Volume 51, Issue 1, pp. 117-122, Feb., 2005.

---

저 자 소 개

---

윤 원 중 (학생회원)

전자공학회논문지 제 42권 SP편 제 3호 참조

박 규 식 (정회원)

전자공학회논문지 제 42권 SP편 제 3호 참조