

다차원 색인구조를 이용한 내용 기반 음악 정보 검색 기법

김 소 영¹⁾ 김 유 성
인하대학교 전자계산공학과
g1991278@inhavision.inha.ac.kr yskim@inha.ac.kr

A Content-based Music Information Retrieval Mechanism Using Multidimensional Indexing Structure

So-Young Kim¹⁾ Yoo-Sung Kim
Dept. of Computer Science & Engineering, Inha University

요 약

오늘날, 컴퓨터 하드웨어와 네트워크의 발달로 인하여, 사용자의 멀티미디어 정보 검색 시스템에 대한 요구들이 높아져가고 있다. 이러한 멀티미디어 정보 검색 시스템에서 멀티미디어 정보는 그 해당 데이터의 고유한 성질에 알맞은 특징 정보로써 표현되며, 각각의 특징 정보를 이용하여 해당 멀티미디어 정보 검색을 위한 색인을 구성하고 사용자의 질의에 대해 검색을 수행하게 된다. 그러나 이미지나 비디오 등의 다른 멀티미디어 정보 검색에 비해, 음악 정보에 대한 검색은 아직 연구가 미비한 상태이므로 본 논문에서는 음악 정보로부터 추출된 특징 정보를 이용하여 음악정보의 다차원 색인을 구축함으로써, 사용자의 음악 질의에 대해서 보다 나은 기능들과 효율을 가지도록 내용 기반 음악 정보 검색 수행을 지원하는 정보 검색 기법을 제안한다.

1. 서 론

내용 기반 멀티미디어 정보 검색 시스템으로 QBIC([1]), Virage Image Search Engine([2]), WebSEEK([3]) 등의 여러 가지 시스템들이 이미 개발되었으며, 아직도 활발한 연구가 진행 중이다. 그러나 이러한 멀티미디어의 내용 기반 검색은 지금까지 이미지나 비디오 검색 분야에 한정되어 연구되어 왔으며, 음악 정보의 경우에는 최근 발표된 몇 가지 내용 기반 음악 정보 검색 시스템이 있을 뿐 아직 초보적인 연구 단계이다. 문제는 실제적으로 사람들은 찾고자 하는 곡에 대하여 그 곡의 메타정보보다는 곡이 가지고 있는 선율의 일부분을 기억하고 질의 작성에 이용한다는 것이다. 따라서 음악 정보의 검색에 있어서 질의의 방법으로 사용자가 기억하고 있는 곡의 일부분을 입력받아서 입력받은 부분의 선율 자체가 가지고 있는 특징 정보를 기반으로 하여 검색을 수행함으로써, 사용자로 하여금 원하는 곡을 검색하는데 있어 편리성을 갖도록 한다.

내용 기반의 음악 정보 검색을 위하여 최근 국내외적으로 음악 정보 검색을 위한 새로운 질의 처리 방법, 선율의 특징과 주제 추출 방법, 그리고 색인 방법의 여러 가지 연구가 진행되고 있다[4,5,6,7]. 그러나 이러한 시스템들은 여전히 그 기능과 성능의 면에서 여러 가지 문제들을 가진다. 따라서 본 논문에서는 많은 음악 정보로 구성된 대용량의 음악 정보 데이터베이스를 효과적으로 관리하고 검색하기 위해, 추출된 음악 정보의 특징 정보를 다차원 색인 기법을 사용하여 검색함으로써, 내용 기반 음악 정보 검색의 성능 및 효율성을 높인다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로써 기존에 제안된 내용 기반의 음악 정보 검색에서의 문제점들을 지적하고, 3장은 본 논문에서 제안하는 음악 정보의 내용 기반 검색을 위한 색인 기법을 전처리 단계와 색인 및 검색의 두 단계로 나누어 살펴본다. 그리고 마지막으로 4장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

본 장에서는 음악 정보에 대한 내용 기반 검색을 제안한 기존의 검색 기법으로 코넬대학에서 1995년에 발표한 Query by Humming (이후 QBH) 시스템([5])과 1997년에 발표한 뉴질랜드의 와이카투(Waikato) 대학의 MELody Index (이후 MELDEX) 시스템([4]) 그리고 1997년에 숭실대에서 발표된 선율을 이용한 음악정보 검색 시스템([6])과 1999년 경북대에서 제안한 리듬유형의 분류에 기반한 효율적인 한국 민요 검색 기법([7])에 대하여 살펴본다.

QBH 시스템과 MELDEX 시스템은 오디오 질의(audio query)를 사용하여 사용자가 원하는 곡을 찾아주는 내용 기반 음악 정보 검색 시스템으로 부정확한 사용자의 흥얼거림으로부터 특징 정보로써 음높이 변화의 방향([8])을 문자열로 나타낸 음의 흐름(contour)을 사용한다. 곡에 대한 음의 흐름은 바로 전의 음과 같은 음높이를 갖는 음이면 R(repeat)(QBH 시스템은 S(same) 사용), 높은 음이면 U(up), 낮은 음이 나오면 D(down)로 표현된다. 따라서 하나의 곡은 R, U, D의 알파벳으로 이루어진 하나의 문자열로 나타내어지는 것이다(그림 1).



그림 1. 동요 '학교중' 첫 악절의 음의 흐름

문자열로 나타낸 특징 정보와 유사한 곡을 검색하기 위해 MELDEX 시스템은 근사 탐색을 위한 상태 일치 알고리즘을 사용해 임의의 k 개의 에러를 허용하여 유사한 곡을 탐색하며, QBH 시스템은 일치되는 것 중 k 개의 에러를 허용하는 퍼지(fuzzy)를 이용한 패턴 매칭 알고리즘인 근사 문자열 탐색(approximate string matching) 알고리즘을 사용하였다.

[6]의 경우, 질의 처리를 위해 음향 기기를 통해 사용자가 흥얼거린 음이나 틀로 만든 곡을 입력으로 받아 곡의 음표로 인

식하고, 그 음표들에 대한 특징 정보로서 문자열을 만든다. 즉, 곡의 음높이를 앞의 음에 대해 세 가지 경우로 나누어 상대적인 음 높이의 1도의 차이를 기준으로 높아지는 경우 A부터 M, 낮아지는 경우 O부터 P까지, 변화가 없는 경우 N으로 나타낸다. 검색을 위해 동적 프로그래밍 알고리즘, 상태 대조 알고리즘 등에서 제안하고 있는 음표열의 근사 문자열 탐색 알고리즘을 사용하여 불일치를 허용하는 문자의 개수를 변화시켜가며 실험하였다. 검색을 위해 첫째 동기로 트라이(trie) 구조의 색인을 구성하여 입력된 특징 정보 문자열의 패턴과의 비교를 통해 탐색하고, 검색이 실패할 경우에는 메타데이터를 이용해 데이터베이스 내의 모든 곡을 대상으로 다시 검색한다.

[7]에서는 일반적인 음악이 아닌 한국 민요의 검색을 위해 음악 정보의 특징을 리듬의 유형으로 선택하여 특징 정보를 추출하였다. 리듬의 유형은 선율에서의 상대적인 음길이 변화의 시퀀스로서 표현되고 이 유형을 인지하기 위해 신경망을 사용하였으며, 검색을 위하여 근사 문자열 탐색 알고리즘을 사용하였다. 이 검색의 방법은 이제까지의 검색 기법에서 특징 추출 시에 소실된 정보인 음의 길이를 고려한 검색이지만 한국 민요라는 특정한 한 장르에 국한된 것이므로 다른 음악 정보의 검색에는 적당하지 않다.

위의 검색 시스템들은 모두 음악 정보에 대하여 내용 기반의 검색을 지원하지만 여전히 다음과 같은 문제점을 가지고 있다.

- 부정확한 특징 정보 추출
- 곡 단위의 비교단위 사용
- 질의의 부정확성을 고려한 검색으로 인한 성능 저하

3. 음악 정보의 색인 기법

본 장에서는 음악 정보에 대한 다차원의 색인을 구성에 대해, 전처리 단계로서 특징 정보와 대표 선율을 추출 방법과 추출된 특징 정보를 다차원 색인 기법 중 하나인 M-tree를 이용한 색인 및 검색하는 방법의 두 단계로 나누어 살펴본다.

3.1 전처리 단계

본 절에서는 먼저 내용 기반 음악 정보 검색의 색인을 위한 전처리 과정에 대해 살펴본다(그림 2).

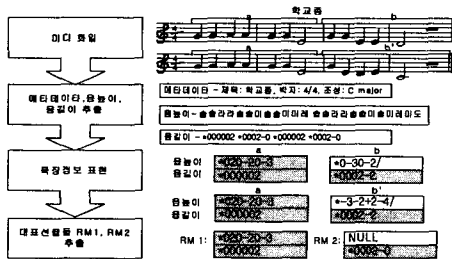


그림 2. 음악 정보의 전처리 과정

먼저 내용 기반 음악정보 검색 시스템에서 사용자는 자신이 원하는 곡을 검색하기 위해 곡의 일부분을 흥얼거린 음악 파일, 특정한 틀을 사용하여 작성한 음악 파일, 메타데이터(곡의 관련 정보), 혹은 음악 화일+메타데이터의 형태로 질의를 던진다. 질의가 되면, 음악 정보 변환의 단계를 거쳐 컴퓨터의 국제적 표준음악화일인 미디(MIDI: Musical Instrument Digital Interface) 화일로 변환한다.

다음으로 변환된 파일로부터 특징 정보를 추출하기 위해 여

러 개의 음을 일정한 규칙에 따라 시간적으로 배치한 것으로 곡의 얼굴이라고 할 수 있는(9) 선율로부터 음높이와 음길이 모두를 정확히 추출한다. 그러면 추출된 정보는 해당 곡을 가장 잘 표현할 수 있는 특징 정보가 된다. 추출 과정은 곡을 음악에서 의미를 갖는 최소 단위인 동기 즉, 두 마디로 분할하여 동기의 단위 별로 수행되며, 추출된 음높이와 음길이에 대하여 조성이나 박자가 바뀌더라도 동일한 곡에 대하여 같은 특징 정보가 추출되어야 한다. 먼저 키 독립성을 유지하는 상대적인 음높이 특징 정보의 추출을 위해, 해당 곡의 첫 번째 나오는 음의 높이를 임의의 '*'로 놓고 그 다음에 나오는 음의 높이를 바로 전에 나온 음의 높이와 비교해 그 차이 값을 받은 단위로 1씩 더하여, 변화의 방향이 올라가는 경우 (+)(생략 가능) 내려가는 경우 (-)로 표현하며, 쉼표의 경우 '/'로 나타낸다. 얻어진 숫자 열은 선율의 곡의 키와 무관하게 상대적인 음높이의 변화 정도를 나타낸다. 이러한 선율의 특징 추출은 기존의 검색 시스템에서 사용된 문자열에 비해 단순히 음높이의 변화 방향뿐만 아니라 방향과 그 변화의 정도까지도 나타낼 수 있다는 장점을 갖는다.

음길이 특징 정보 추출의 경우는 첫 번째 나타나는 음의 길이를 '*'로, 그 다음에 나오는 음의 길이를 바로 전에 나온 음의 길이와 비교하여 다음에 나오는 음표의 길이가 바로 전에 나온 음표의 길이보다 길어지거나 짧아지는 것을 비율로서 나타나는 숫자 열로서 표현하며, 쉼표의 경우는 해당 음표의 상대적인 길이의 앞에 (-)를 붙인다. 이렇게 표현된 숫자 열은 사용자가 동일한 곡을 다른 빠르기로 부르더라도 동일한 곡으로 인지할 수 있도록 하는 템포 독립성(tempo independency)을 가진다. 따라서 기존의 내용 기반 시스템에서 선율의 특징 정보 추출과정에서 음길이 정보를 전부 상실하였던 것과 비교하여 훨씬 정확한 특징 정보를 추출한다.

마지막으로 추출된 특징 정보로부터 대표 선율을 결정한다. 주제(theme)가 되는 선율은 곡의 전체 분위기를 결정짓는 2마디, 4마디, 혹은 8마디로 곡에서 한번 이상 반복이 되어 듣는 사람에게 곡에 통일감을 주며, 주제를 인식할 수 있도록 한다 [9]. 따라서 주제는 곡 전체에서 가장 많은 반복되는 부분이라고 다시 정의되어질 수 있다. 결정은 곡의 선율에서 가장 높은 음이 나오는 부분이며, 주제선율과 일치할 수도 그렇지 않을 수도 있다. 그림 3에서는 그림 2에서 선율의 음높이와 음길이 유형을 기반으로 대표 선율 추출의 예를 보인다.

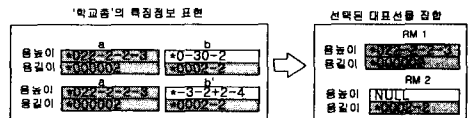


그림 3. '학교중'의 대표 선율 추출

3.2 음악 정보의 색인기법

본 절에서는 먼저 음악 정보 객체에 대해 각각의 특징 정보의 대표선율을 검색을 위한 키워드들로서 다차원의 공간 내에 표현하고, 이 특징 정보들간에 서로의 유사도를 정의할 수 있는 계산식을 이용하여, 계산된 유사도를 기준으로 그 유사한 정도가 어떤 기준 이내에 존재하는 모든 객체들을 검색하는 유사질의를 수행하기 위한 다차원 색인 기법인 M-tree[10,11,12]를 이용한 색인 기법을 제안한다.

먼저 매트릭 공간은 한 쌍으로, $M=(D, d)$ 의 식으로 나타내어지며, 여기서 D는 특징 정보의 도메인 값(색인 키 값)으로 본 논문에서는 (음높이 변화의 유형, 음길이 변화의 유형)의 형식

으로 표현되며, 거리의 함수 d는 음악 정보가 가지는 특징 정보들 간의 유사도를 계산하는 거리의 함수로써 다음의 식 1로 표현할 수 있다.

$$d = w_1 \times \left(\sum_{i=1}^n (P1_i - P2_i)^2 \right) + w_2 \times \left(\sum_{i=1}^n (L1_i - L2_i)^2 \right) \quad \text{식 1}$$

- $P1_i, P2_i$: 주제선을 내의 음높이들의 상대적 변화 값
- $L1_i, L2_i$: 주제선을 내의 음길이들의 상대적 변화 값
- w_1 : 음높이의 가중치
- w_2 : 음길이의 가중치
- n : 정규화된 음표의 개수

위의 식 1은 M-tree 구성 시, 거리의 함수가 가져야하는 대칭성(symmetry), 절대성(positivity), 그리고 삼각 부등식(triangle inequality)을 만족한다. 음악 정보의 유사질의 처리를 위한 M-트리 알고리즘은 접근해야하는 노드의 수와 거리 계산의 횟수를 줄이기 위한 것으로 사용자로부터의 불명확한 질의를 고려하기 위해 어느 정도의 거리 즉, 비유사도를 허용하여 그 범위 내에 존재하는 모든 곡을 검색하는 범위 질의를 수행한다. 즉, 음악 정보 검색을 위한 범위 질의 탐색 알고리즘(range search algorithm: 이후 RS)은 루트 노드로부터 시작해 반복적으로 부등식을 만족하는 모든 경로를 반복적으로 순회하여 사용자가 질의한 음악 정보 M_q 와 데이터베이스 내의 음악 정보 M_d 간의 유사도를 측정하여 허용 범위 내에 있는 음악 정보들을 검색하게 된다. 여기서 허용 범위는 $r(M_q)$ 가 된다. 이와 같은 내용 기반 음악 정보 검색을 위한 범위 질의 탐색 알고리즘을 알고리즘 1에서 나타내었다.

(알고리즘 1) 음악 정보의 범위 질의 탐색 알고리즘

```

RS(N: node, Q: query object, r(Q): search_radius) {
  if (N == root node) then {
    for VO, in N do {
      if (d(O, Q) ≤ r(Q) + r(O)) then
        RS(*ptr(T(O)), Q, r(Q));)
    }
  }
  else {
    let Op be the parent object of N;
    if (N != leaf node) then {
      for VO, in N do {
        if (|d(Op, Q) - d(Op, Op)| ≤ r(Q) + r(Op)) then {
          RS(*ptr(T(Op)), Q, r(Q));)
        }
        else {
          for VO, in N do {
            if (|d(Op, Q) - d(Op, Op)| ≤ r(Q)) then
              add oid(Op) to the result; )))
        }
      }
    }
  }
} /* Range search algorithm */
    
```

명제 1. 만약 $d(O, Q) > r(Q) + r(O)$ 이면, $T(O)$ 의 각 객체 O_p 에 대해 $d(O, Q) > r(Q)$ 이다. 따라서 $T(O)$ 은 안전하게 검색에서 제외될 수 있다.

사실 $d(O, Q) > d(O, Q) - d(O, O_p)$ 이고 $d(O, O_p) ≤ r(O)$ 이므로, $d(O, Q) > d(O, Q) - r(O)$ 이다. 가설에 의해 $d(O, Q) > r(Q)$ 이다. 명제 1을 적용하기 위해서는 $d(O, Q)$ 을 계산해야 하지만 다음의 명제 2에 따라 계산의 복잡성을 피할 수 있다.

명제 2. 만약 $|d(O_p, Q) - d(O, O_p)| > r(Q) + r(O)$ 이면, $d(O, Q) > r(Q) + r(O)$ 이다.

이것은 $d(O, Q) ≥ d(O_p, Q) - d(O, O_p)$ 와 $d(O, Q) ≥ d(O, O_p) - d(O_p, Q)$ 를 보장하는 삼각 부등식에 의해 유지된다. 동일한 이론이 일노드에도 적용되며, 실험 결과 이 기술은 40%의 거리 계산을 절약할 수 있음을 보였다[12].

4. 결론 및 향후 연구 방향

본 논문에서 제안한 음악 정보의 내용 기반 검색을 위한 색인기법은 많은 양의 음악 정보를 데이터베이스에 저장하고 효율적으로 검색할 수 있도록 설계된 알고리즘이다. 기존에 제안된 음악 정보의 내용 기반 검색기법과 비교하여 특징 정보의 추출과정에서 주제선율의 특징 정보 중, 음의 길이와 음의 높이의 정보를 모두 포함함으로써, 초기 검색의 정확율을 높이고 다차원 색인기법으로 주제선율 또는 절정을 주제선율 색인하여 음악의 특징 정보 데이터베이스를 구축함으로써, 기존의 시스템들의 검색 비교단위가 곡 전체였던 것에 비해 공간적 낭비와 탐색 시간의 지연을 개선하였다.

향후 연구로서 본 논문에서는 제안된 음악 정보를 위한 색인기법을 이용하여 음악 정보 내용 기반 검색 시스템을 구현하고, 그 타당성을 검증하는 것이 필요하다.

참고문헌

- [1] W. Niblack, et al., "QBIC Project: Querying Images by Content, Using Color, Texture, and Shape," *Proceedings of the Conference on Storage and Retrieval for Image and Video Databases*, Feb.2-3, 1993, Workshop, 1999.
- [2] J. R. Bach, "The Virage Image Search Engine: An Open Framework for Image Management," *Proceedings of the Conference on Storage and Retrieval for Image and Video Databases*, Feb. 1-2, 1996.
- [3] J. R. Smith, and S. F. Chang, "Tools and Techniques for Color Image Retrieval," *Proceedings of the Conference on Storage and Retrieval for Image and Video Databases*, Feb. 1-2, 1996.
- [4] R. J. McNab, L.A. Smith, I. H. Witten, C. L. Henderson and S.J. Cunningham, "Towards the Digital Music Library: Tune Retrieval from Acoustic Input," *Digital Libraries*, 1996.
- [5] A. Ghias, J. Logan, D. Chamberlin and B.C. Smith, "Query By Humming Musical Information Retrieval in an Audio Database," *ACM Multimedia 95*, November 1995.
- [6] Jeong-Gyu Jee, Heysock Oh, "Design and Implementation of Music Information Retrieval System using Melodies," *The Korea Information Processing Society*, Vol.5, No.1, 1998.
- [7] J. T. Shin, J. W. Kim, H. J. Kim, and P.W. Park, "Efficient Korean Folk Song Retrieval Method based on Rhythm Pattern Classification," *ITC-CSCC*, 1999.
- [8] W. J. Dowling, "Scale and Contour: Two components of a theory of memory for melodies," *Psychological Review*, 1978.
- [9] 이병욱, 백기풍, "세 번만 읽으면 누구나 작곡할 수 있다," *Jackeunwoori Pub. Co.*, 1989
- [10] P. Zezula, P. Ciaccia, and F. Rabitti, "M-tree: a dynamic index for similarity queries in multimedia databases," *Technical Report, HERMES ESPRIT LTR Project*, 1996.
- [11] P. Ciaccia, M. Patella, F. Rabitti, P. Zezula, "Indexing Metric Spaces with M-tree," *ACM Transaction*, 1999.
- [12] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An Efficient Access Method for Similarity Search in Metric Spaces," *Proceedings of the 23rd VLDB*, 1997.
- [13] Jong-Sik Mo, Chang-Ho Han, and Yoo-Sung Kim, "A Melody-based Similarity Computation for Musical Information," *Proceedings of the Workshop on Knowledge and Data Engineering Exchange*, 1999.