

Hierarchical ART2 알고리즘을 이용한 악보 영상 인식

김미정^{*} · 김재균^{**} · 박충식^{***} · 김광백^{*}

^{*}신라대학교 컴퓨터정보공학부

^{**}신라대학교 컴퓨터공학과

^{***}영동대학교 컴퓨터공학과

Music Image Recognition using Hierarchical ART2 Algorithm

Mi-jeong Kim^{*} · Jae-kun Kim^{**} · Choong-shik Park^{***} · Kwang-baek Kim^{*}

^{*}Division of Computer and Information Engineering, Silla University

^{**}Dept. of Computer Engineering, Silla University

^{***}Dept. of Computer Engineering, Youngdong University

요 약

음악 연구에 따른 컴퓨터의 역할이 점차 중요한 비중을 차지함에 따라 보다 효과적인 악보 인식과 효율적인 악보의 편집 및 수정 방법이 요구된다. 기존의 수동 입력 방식에서는 악보를 부정확하게 입력하여 수정하는 경우에는 작업 시간이 많이 소요되며, 각 수정 프로그램에서 만든 악보는 특정 프로그램에서만 재수정이 가능하다는 단점이 있다. 본 논문에서는 이러한 단점을 보완하기 위하여 이미 작성 되어있는 악보들을 자동으로 인식하는 방법을 제안한다. 제안된 악보 인식 방법은 수평 히스토그램을 이용하여 악보 이미지의 오선을 제거한 후, 4방향 윤곽선 추적 알고리즘을 적용하여 잡음을 제거하고 Grassfire 알고리즘을 적용하여 악보 구성 기호들을 추출한다. 추출된 악보 구성 기호들은 Hierarchical ART2 알고리즘을 적용하여 인식한다. 인식된 악보 구성 기호들을 이용하여 악보 구성 기호들이 속하는 마디의 위치 정보를 각각 저장하고 향후에 악보 구성 기호의 편집과 수정이 용이하게 한다. 제안된 악보 인식 방법의 성능을 평가하기 위해 100장의 악보 영상을 대상으로 실험한 결과, 제시된 Hierarchical ART2 알고리즘을 이용한 악보 영상의 인식 방법이 실험을 통해서 효율적인 것을 확인하였다.

키워드

히스토그램, Hierarchical ART2, 악보 구성 기호, 세션화

1. 서 론

최근 디지털 정보에 대한 많은 사용으로 인해 신속하고 정확한 정보가 요구되고 있다. 이것은 IT 분야뿐만 아니라 음악과 같은 예능 분야에서도 요구된다. 따라서 악보 작성이 디지털화됨으로 인하여 악보가 전용 소프트웨어로 제작되고 있지만, 기존의 악보 제작 소프트웨어들은 다른 소프트웨어들로 작성된 악보들은 편집하

지 못하거나 인쇄된 악보의 경우에는 편집이 어렵다는 단점이 있다. 악보의 편집을 위해 수작업으로 악보를 입력하게 되면 빠르고 정확한 입력을 할 수가 없다[1]. 따라서 본 논문에서는 이러한 문제점을 개선하기 위하여 스캔으로 획득한 악보 영상에 대한 인식 방법을 제안한다. 악보의 정확한 인식을 위해 획득한 악보 영상에 평균 이진화를 적용하여 이진화 하고, 이진화된 영상을 4방향 윤곽선 추적 알고리즘을 적용하여 악보 구성 기호들의 잡음 영역을 제거

한다. 수평 히스토그램을 이용하여 오선을 제거하고, 세선화 알고리즘을 적용하여 객체의 골격을 추출한 후에 Grassfire 알고리즘으로 모든 악보 구성 기호를 찾는다. 찾은 객체의 수직 히스토그램을 이용하여 잇단 음표를 잘라내고, 악보구성기호의 크기를 정규화한 후에 Hierarchical ART2 알고리즘을 적용하여 악보 구성 기호를 인식한다.

II. 악보 구성 기호 추출

본 논문에서 제안한 악보 구성 기호 추출 방법은 평균 이진화 기법을 적용하여 악보 영상을 이진화 한 후에 4방향 윤곽선 추적 알고리즘으로 작은 악절 영역을 추출한다. 추출된 작은 악절 영역의 수평 히스토그램을 이용하여 오선을 제거한 후, Grassfire 알고리즘을 이용하여 악보 구성 기호들을 추출하고 서로 연결되어 있는 음표들을 분리한다. 그림 1은 제안된 악보 구성 기호를 추출하는 과정을 나타낸 것이다.

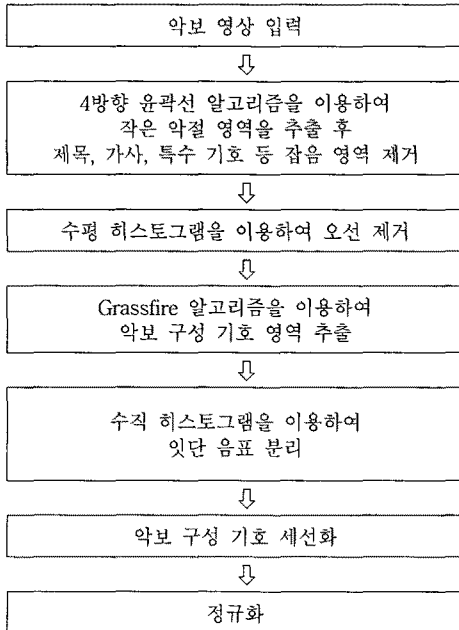


그림 1. 악보 구성 기호 추출 과정

그림 2와 같이 획득된 악보 영상에서 악보 구성 기호들을 추출하기 위해, 영상 안에 모든 픽셀에 대한 평균값을 임계치로 하는 평균 이진화 기법을 적용하여 영상을 이진화 한다. 이

진화 된 영상에 4방향 윤곽선 추적 알고리즘을 적용하여 작은 악절 영역을 추출한다.



그림 2. 입력 된 악보 영상(좌)
평균 이진화 된 악보 영상(우)

작은 악절은 음의 높이를 정확하게 나타내기 위해 사용되는 오선과 음의 자리를 정하기 위한 음자리표, 그리고 각종 음표와 기호들로 구성되어 있다. 음표의 구성 요소 중 음표의 기둥이 없는 온음표는 높은음자리표의 높이를 넘지 않고, 음표의 기둥이 있는 4분음표와 8분음표의 경우에는 그림 3과 같이 음표의 머리가 오선에서 4마디 위 아래로 까지 내려가도 오선과 연결되는 특징이 있다. 그리고 이진화 된 영상에는 오선 및 제목, 노랫말, 특수기호 등과 같은 기타 잡음 영역이 포함되어 있기 때문에 이러한 영역을 제거하여 악보 구성 기호를 정확하게 추출한다.

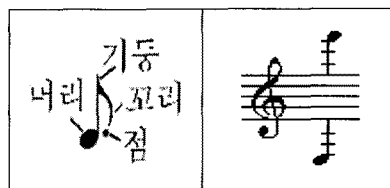


그림 3. 음표의 구성과 오선의 범위

그러나 오선 영역은 악보 구성 기호 영역과 함께 추출되기 때문에, 오선과 악보 구성 기호들의 영역을 분할하여 제거한다. 오선은 작은 악절의 수평 히스토그램을 이용하여 오선의 위치를 파악하고, 오선과 악보 구성 요소와의 교차점에 있는 상하 픽셀(Pixel)을 측정하여 흑화소가 없을 경우에는 오선으로 간주하여 제거한다[2,3]. 그림 4는 추출한 작은 악절 영역과 오선 및 기타 잡음 영역을 제거한 영상이다.



그림 4. 작은 악절 영역을 추출한 악보 영상(좌) 오선과 잡음이 제거된 악보 영상(우)

악보 구성 기호들은 오선과 기타 잡음 영역들이 제거된 영상에 Grassfire 알고리즘을 적용하여 추출한다. Grassfire 알고리즘은 자기호출을 이용하여 모든 인접 요소가 라벨링될 때까지 현재 관심 화소의 주변 인접 화소를 차례로 검사하면서 라벨링하는 방법이다. 그림 5는 Grassfire 알고리즘을 적용하여 악보 구성 기호를 추출한 결과이다.

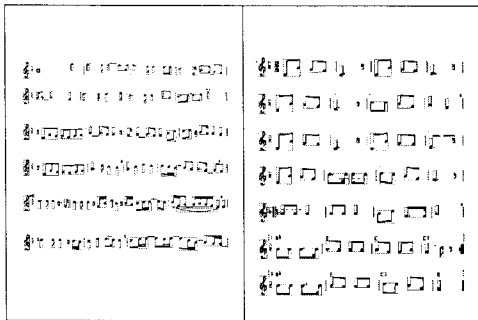


그림 5. Grassfire 알고리즘을 적용하여 악보 구성 기호들을 추출한 영상

악보 구성 요소는 각각의 다른 영역 비율로 인해 정규화시 악보 구성 요소의 두께가 변하는 문제점이 있어 정규화 과정에서 비율에 대한 오차를 최소화하기 위해 Grassfire 알고리즘을 적용한 영상에 세선화 방법[4]을 적용한다. 세선화는 영상의 골격만을 추출하는 영상처리 기법으로, 여러 픽셀로 된 두꺼운 영역을 1 픽셀 값의 일률적 선으로 표시하여 두꺼운 선들의 골격만을 추출하는 영상처리 기법이다. 악보 구성 기호를 세선화한 영상은 그림 6과 같다.

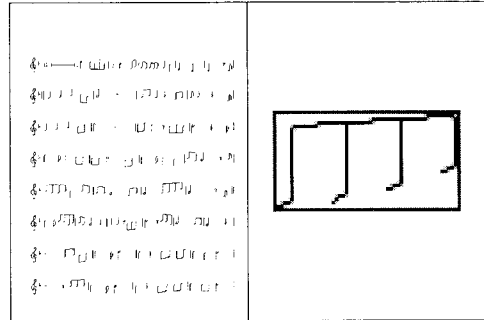


그림 6. 세선화한 영상

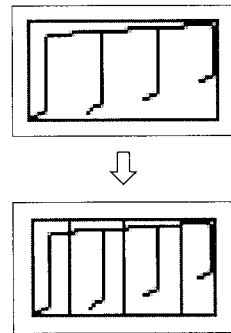


그림 7. 잇단 음표 분리

세선화된 악보 구성 기호들 중에서 그림 7과 같이 꼬리가 연결되어 있는 잇단 음표들은 수직 히스토그램을 이용하여 기둥 영역을 찾은 후에 영역을 분할한다. 영역 분할이 완료된 악보 구성 기호들을 Hierarchical ART2 알고리즘의 입력 패턴으로 적용하기 위해 영역을 일정한 크기로 정규화 한다. 그림 8은 악보 구성 기호를 정규화하여 나타낸 것이다.

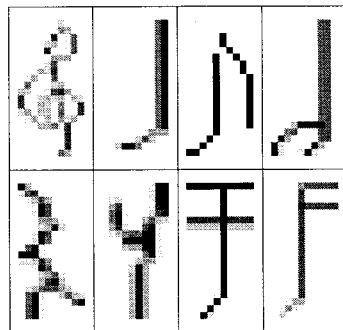


그림 8. 세선화 된 악보 구성 기호의 정규화

III. Hierarchical ART2 알고리즘을 이용한 악보 구성 기호 인식

본 논문에서는 추출된 악보 구성 기호의 학습 및 인식을 위해 Hierarchical ART2 알고리즘을 적용한다. ART2 신경회로망은 기존에 학습되어있었던 것이 새로운 학습에 의해 지워지지 않도록 새로운 지식을 자동으로 전체 구조에 일관성 있는 방법으로 통합하며 실시간으로 빠르고 안정적으로 학습할 수 있는 적응적인 구조를 가지고 있다.

ART2 신경회로망의 구조는 경계 변수 ρ 에 의해 결정된다. 여기서, ρ 는 임의의 입력 패턴과 저장된 클러스터 패턴과의 일치 허용도로 정의한다. 즉, ρ 값이 크게 되면 클러스터의 반경이 커지게 됨으로써 하나의 클러스터가 많은 패턴을 수용하게 되어 클러스터의 개수는 작아지는 반면 다양한 패턴들을 저장할 수 없게 된다. 반대로, ρ 값이 작으면 클러스터의 반경이 작아지므로 많은 클러스터가 생성되어 다양한 패턴들을 저장할 수 있지만 학습 및 인식에 있어서 비교해야할 클러스터의 개수가 너무 많기 때문에 속도의 저하를 가져온다.

본 논문에서는 이러한 경계 변수의 특성을 고려하여 두 개의 층으로 구성된 클러스터링 방법인 Hierarchical ART2 알고리즘[5]을 적용하여 인식한다.

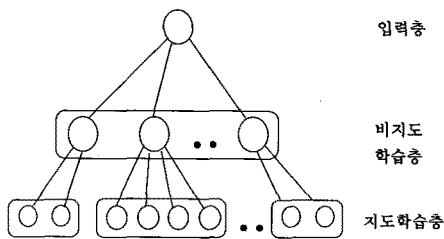


그림 9. H-ART2 신경회로망 구조

Hierarchical ART2는 두 개의 층으로 된 클러스터로 구성되어 있으며 입력층과 연결된 1차 클러스터와 클러스터 각각에 연결된 2차 클러스터들로 구성되어 있다. 여기서 1차 클러스터는 기존의 ART2 학습 알고리즘에 의해 생성되는 중심 참조 벡터 패턴(centroid reference vector pattern)으로써 입력 패턴 집합을 대표하는 대표 패턴이다. Hierarchical ART2는 1차 클러스터를 통해 입력 패턴의 형태를 대략적으로

분류하고 각각의 1차 클러스터에 연결된 2차 클러스터를 통해 다양한 형태의 패턴을 대표하게 된다.

H-ART2의 1차 클러스터의 생성은 기존의 ART2 방법과 동일한 방법을 사용하며 비지도 학습으로 생성된다. 여기서 생성되는 클러스터는 대략적으로 입력 패턴의 유형을 분류하게 된다. 각각의 1차 클러스터는 거기에 속한 입력 패턴들을 다시 2차 클러스터로 분류하는데 이때에는 ART2 학습 알고리즘을 지도학습으로 변형하여 세부적으로 분류한다. ART2의 지도 학습 알고리즘은 비지도 학습과 유사하지만 유사성 검증을 통과한 입력 패턴이라 할지라도 입력 패턴이 승자 클러스터와 다른 클래스일 경우에는 새로운 클러스터를 생성한다는 점이 다르다. ART2 지도학습 알고리즘은 다음과 같다.

단계 1: k번째 입력패턴을 x_k , 신경회로망의 i번째 클러스터의 중심값을 w_i 이라 하고 LX_k , LC_i 를 각각 입력 패턴 x_k 와 클러스터 중심값 w_i 의 해당 클래스를 나타내는 레벨로 정의한다.

단계 2: 새로운 입력 패턴 x_k 에 대해 최소 거리를 가지는 클러스터 j^* 을 승자 클러스터로 선택한다.

단계 3: 입력 패턴에 대한 유사성 검증을 수행한다. 만약 입력 패턴이 승자 클러스터의 중심 반경 (Vigilance Parameter) 이내에 있고 입력 패턴과 승자 클러스터가 같은 클래스일 경우에는 이 입력 패턴을 승자 클러스터에 포함시키고 그 클러스터의 중심값을 조정한다. 그렇지 않은 경우에는 현재 입력 패턴 x_k 를 새로운 클러스터로 할당한다.

$$w_{j^*}^{new} = \frac{x_k + w_{j^*}^{old} \cdot \|Cluster_{j^*}^{old}\|}{\|Cluster_{j^*}^{old}\| + 1} \quad (1)$$

단계 4: 모든 입력이 제시될 때까지 (1)~(3)의 과정을 반복 수행한다.

단계 5: 신경망의 클러스터 중심값이 각각 변함이 없으면 학습을 종료한다.

IV. 실험 및 결과 분석

실험 환경은 Intel Pentium-IV 2GHz CPU와 512MB RAM이 장착된 IBM 호환 PC상에서 VC++ 6.0으로 구현하였다. 제안된 악보 인식 방법의 성능을 평가하기 위해 100장의 악보 영상을 스캔하여 획득된 악보 영상을 대상으로 실험하여 인식 성능을 평가하였다.

본 논문에서 제안한 악보 영상 추출 및 인식 방법을 적용한 결과, 표 1에서와 같이 100장의 악보 영상에서 10734의 악보 구성 기호 중에서 10467개의 악보 구성 기호가 추출되었다.

표 1. 악보 구성 기호 추출 결과

	제안된 방법 (추출 개수 / 총 개수)
악보 구성 기호 코드	10467 / 10734

개별 코드 인식을 위해 적용된 Hierarchical ART2 알고리즘의 학습 및 인식 성능을 평가하기 위하여 100개의 악보 영상에서 추출한 악보 구성 기호 88개를 학습 패턴으로 적용하였다. 그리고 테스트 패턴은 기호 코드 10467개를 테스트 하였다. Hierarchical ART2 알고리즘으로 인식한 결과는 표 2와 같다. 표 2에서와 같이 음표 코드는 8876개 중에서 8165개가 인식되었고 쉼표 코드는 741개 중에서 596개가 인식되었다. 높은음자리표나 올림표, 내림표 등의 나머지 코드들은 995개 중 980개가 인식 되었다. 따라서 Hierarchical ART2 알고리즘이 개별 코드 인식에 있어서 효율적인 것을 확인할 수 있다. Hierarchical ART2 알고리즘으로 악보 구성 기호의 인식에 실패한 경우는 추출한 개별 문자에 잡음이 제거되지 않은 경우와 음표와 쉼표가 함께 추출된 경우이다. 그림 10은 개별 문자 인식에 실패한 악보 구성 기호를 나타내었다.

표 2. 악보 구성 기호 인식 결과

	음표	쉼표	그 외 코드	합계
인식개수	8165 / 8876	523 / 596	980 / 995	9668 / 10467



그림 10. 인식에 실패한 악보 구성 기호

VI. 결 론

본 논문에서는 악보 구성 기호를 추출하고 인식하는 방법을 제안하였다. 악보 구성 기호의 추출은 악보 영상을 그레이 레벨로 변환한 후, 평균 이진화 방법을 적용하여 이진화하고 4 방향 윤곽선 알고리즘을 적용하여 작은 악절 영역을 추출하였다. 작은 악절 영역에 포함되지 않는 제목, 가사 등은 잡음으로 간주하여 제거하였고, 작은 악절 영역에서 오선은 수평 히스토그램을 적용하여 제거하였다. 악보 구성 기호를 추출하기 위해 Grassfire 알고리즘으로 라벨링 하여 악보 구성 기호를 추출하였고, 추출된 악보 구성 기호들을 확대할 경우에 발생하는 두께의 차이를 줄이기 위하여 세선화 하였다. 세선화 된 기호들 중 잇단 음표들은 수직 히스토그램을 적용하여 분리시키고 정규화하여 악보 구성 기호들을 추출하였다. 추출된 기호들은 Hierarchical ART2 알고리즘을 적용하여 각 악보 구성 기호들을 인식하였다.

제안된 악보 구성 기호 추출 및 인식 방법의 성능을 평가하기 위해 악보 영상 100장을 대상으로 실험한 결과, 100장의 악보 영상에 대해 10734개 중에서 10467개가 추출되었다. 추출된 개별 코드를 Hierarchical ART2 알고리즘을 적용하여 인식한 결과 10467개의 악보 구성 기호 중에서 9668개가 인식되었다.

향후 연구 방향은 악보 구성 기호의 추출물을 개선하기 위해 퍼지 마스크를 이용한 방법에 대해 연구할 것이고 현재 적용된 Hierarchical ART2 알고리즘의 문제점을 개선하여 인식률을 향상시킬 것이다.

참고문헌

- [1] 박충식, 장경식, 전정우, 김재희 등, “악보인식을 위한 전처리 과정,” 인공지능, 신경망 및 퍼지관련 학술발표회 논문집, pp.169-175, 1993.
- [2] 조형제, 조경은, “골격선과 런 길이 정보를

- 이용한 피아노 악보 인식,” 정보과학회논문지, 제2권, 4호, pp.461-473, 1996.
- [3] 손화정, 김수형, 오성열, “카메라 기반 악보 영상 인식을 위한 오선 검출 및 삭제 알고리즘,” 한국콘텐츠학회논문지, 제 7권 11호, pp. 34-42, 2007.
- [4] 성기종, 유근호, 이철희, “이진영상의 새로운 세션화 알고리즘,” 대한전자공학회 학술대회 논문집, pp.289-291, 1985.
- [5] 김도현, 차의영, “적응적 탐색 전략을 갖춘 계층적 ART2 분류 모델,” 정보과학회논문지, 제30권, 7·8호, pp.649-658, 2003.