
SOM과 개선된 ART-1을 이용한 악보 인식

김광백*

Musical Score Recognition with SOM and Enhanced ART-1

Kwang-baek Kim*

요 약

본 논문에서는 SOM과 개선된 ART-1을 이용하여 악보를 인식하는 방법을 제안한다. 악보 인식을 위해 스캔된 악보 이미지를 호프 변환, Otsu's 이진화를 원본 이미지에 적용하고, 히스토그램 분석을 통해 구분된 작은악절에서 오선을 제거하여 악보의 음표 성분을 추출할 수 있는 이미지 전처리 단계를 수행한다. 오선이 제거된 작은악절은 레이블링을 이용하여 음표 성분을 분리한다. 추출된 음표들은 SOM 알고리즘을 적용하여 일정한 크기로 정규화하고, 정규화된 음표 정보들을 개선된 ART-1 알고리즘을 적용하여 학습과 인식한다. 제안된 방법을 적용하여 음표 인식 실험을 한 결과, 제안된 방법이 음표 인식에 효율적임을 확인하였다.

ABSTRACT

In this paper, we propose a Musical Score Recognition with SOM and Enhanced ART-1 Algorithm. First, we apply Hough transform and Otsu's binarization to the original BMP format image and extract notes from separated passages by histogram analysis with removing staff lines. Then extracted musical notes are normalized to same size by SOM algorithm and ART-1 algorithm plays the learning and recognition role. The experiment verifies that the proposed method is quite effective on printed musical score recognition.

키워드

SOM, ART-1, 악보 인식, 오선, 음표

Key word

SOM, ART-1, Musical Score Recognition, Staff Lines, Musical Notes

* 종신회원 : 신라대학교 컴퓨터공학과(gbkim@silla.ac.kr)

접수일자 : 2013. 02. 22

심사완료일자 : 2013. 03. 12

I. 서론

최근 영상 인식 분야는 OA 관련 분야, 의료용 영상 처리, 위성 사진 처리, 산업 자동화 분야, 방송 및 영화 관련 엔터테인먼트 산업, VOD를 위한 영상 압축, 개인 사용자를 위한 영상 인식 인터페이스 등 현대 산업 전반에 걸쳐 광범위하게 활용되고 있다[1].

악보 인식을 위한 연구는 1960년대부터 현재까지 활발히 이루어지고 있지만 특정 악보 형태에서만 인식률이 높은 특성을 보이고 있다[2]. 또한 음악 관련 소프트웨어의 경우에 서로 호환이 되지 않거나 인쇄 악보의 경우에는 인식 작업이 어렵다는 단점이 존재한다[3,4]. 따라서 본 논문에서는 인쇄 악보를 인식하기 위해 SOM 알고리즘과[5] 개선된 ART-1 알고리즘을[6] 이용하여 악보의 음표를 인식하는 방법을 제안한다.

기존의 신경망을 사용한 인식 방법들은 Hierarchical ART2 알고리즘을 적용한 방법[4], 오류 역전파 알고리즘을 적용한 방법[2] 등이 존재한다. 하지만 신경망을 이용한 인식 기술은 인식 대상을 일정한 크기로 정규화해야 하는 문제점이 있다. 이를 위해 추출된 음표를 세선화하거나[4], 오선 간격에 따라 정규화 크기를 고정하는 방법[2]을 사용하였으나 인식률에 있어 큰 성과를 거두지 못한 것이 사실이다. 또한 신경망 구조는 알고리즘이 복잡해질수록 필요한 변수와 임시 저장 공간이 늘어나게 되며 시스템의 처리를 느리게 하는 요인으로 작용한다. 본 논문에서는 이미지의 정규화로 인해 발생하는 데이터의 손실을 최소화하기 위해 SOM 알고리즘을 적용하여 이미지 좌표 정보를 고정된 크기의 Map에 매칭시켜 일정한 크기로 정규화하고, 정규화한 정보를 개선된 ART-1 알고리즘을 적용하여 학습 및 인식에 활용하는 방법을 제안한다.

II. SOM과 개선된 ART-1을 이용한 악보 인식

본 논문에서 제안한 인식 방법을 적용하기 위해 기호 추출을 위한 영상의 전처리 단계가 필요하다. 악보 영상은 BMP 포맷의 스캔 영상이다. 영상이 입력되면 호프 변환을 적용하여 영상의 기울어진 정도를 분석한다. 호프 변환은 N차원 좌표상의 점을 거리(r)와 각도

(θ)로 변환하여 직선 또는 방정식으로 이루어진 특정 도형을 검색할 때 사용되는 알고리즘이다[7]. 따라서 호프 변환을 사용하면 문서상에 존재하는 직선들을 검출할 수 있게 되며, 검출된 직선들 중 가장 긴 직선은 오선 중 하나가 된다. 이러한 특징을 이용하여 문서의 기울어진 정도를 분석하여 영상의 평형을 맞춘다. 평형이 맞추어진 영상을 Otsu's 이진화를 사용하여 이진화 과정을 수행한다. 그림 1은 호프 변환과 Otsu's 이진화를 수행한 결과이다.

그림 1. 변환된 이미지
Fig. 1 Transformed Image

이진화된 영상은 오선 탐색을 효율적으로 수행하기 위해 수평 히스토그램을 분석하여 각 악절로 분할하게 된다. 추출된 각 악절은 오선을 제거하여 음표 성분만 남긴다. 그림 2는 오선이 제거된 악보 영상이다. 오선이 제거된 각 악절은 레이블링을 수행하여 기호를 추출한다. 기호 추출 시 잡음과 구분하기 위하여 일정 픽셀 수 이상 레이블링이 되지 않으면 잡음으로 간주하여 기호로 추출하지 않는다.

SOM은 1982년 Kohonen 교수에 의하여 제안된 신경망 알고리즘이다. Kohonen Neural Network(KNN)라고 불리는 SOM 알고리즘은 입력층과 Kohonen층으로 이루어져 있으며 Kohonen층의 각 노드들이 서로 경쟁하여 입력 벡터를 조직화 하는 자율 학습 신경망이다[5].



그림 2. 오선 제거된 영상
Fig. 2 Image after Removing Staff Lines

KNN은 Euclidean 거리 계산법으로 각 입력 벡터로부터 Kohonen층의 각 노드에 대한 거리 값을 계산하여 최소 거리에 존재하는 노드를 승자 노드로 출력하게 된다. 식 (1)은 승자 노드를 출력하기 위한 Euclidean 거리 계산 식이다.

$$Dist_{x,y}(t) = \sqrt{\sum_{k=0}^{\lfloor pl \rfloor} (n_{0,k}(t) - w_{0,k \rightarrow 1,(x,y)}(t))^2} \quad (1)$$

식 (1)에서 (x,y) 는 Kohonen층의 노드 좌표이며 $n_{0,k}$ 은 입력 벡터, $w_{0,k \rightarrow 1,(x,y)}$ 는 각 입력 벡터의 노드에서 Kohonen층의 (x,y) 노드로의 가중치 값이다. 단 본 논문에서는 최소 거리에 대한 승자 노드 좌표 (x,y) 를 요구하므로 제공된 연산은 수행하지 않는다.

식 (1)에 의해서 승자 노드가 도출되면 가중치 값을 변경하게 된다. 식 (2), (3), (4)은 가중치 값을 변경하는 식이다.

$$w_{0,k \rightarrow 1,(i,j)}(t+1) = w_{0,k \rightarrow 1,(i,j)}(t) + \Delta w_{0,k \rightarrow 1,(i,j)}(t+1) \quad (2)$$

$$\Delta = \text{Neigh}_{win(row,col)}(t+1)\beta(t)(n_{0,k}(t) - w_{0,k \rightarrow 1,(i,j)}(t)) \quad (3)$$

$$\begin{aligned} &\text{if}((W_{row*}(t) - H(t)) \leq i \leq W_{row*}(t) + H(t)) \text{ and} \\ &(W_{col*}(t) - H(t)) \leq j \leq W_{col*}(t) + H(t) \\ &N_{i,j}(t) = 1 \\ &\text{else } N_{i,j}(t) = 0 \end{aligned} \quad (4)$$

식 (3)에서 β 는 학습률이고 식 (4)에서 가중치의 변경은 승자 노드를 중심으로 이웃 노드에게만 적용된다. 식

(4)에서 H 는 이웃 반경 값이며 반복 횟수 t 에 따라 변화한다. 반복 횟수 t 에 따라 변화하는 학습률 β 와 이웃반경 H 를 구하는 식은 식 (5), (6)과 같다.

$$\beta(t) = \beta_{init} - \frac{t}{Q}(\beta_{init} - \beta_{fin}) \quad (5)$$

$$H(t) = H_{init} - ((t \div \|\mathbb{A}\|) \div H_{inter}) \quad (6)$$

식 (5)에서 Q 는 학습에 제시된 최대 반복 횟수이며 β_{init} 는 초기 학습률, β_{fin} 는 최종 학습률이다. 식 (6)의 H_{init} 는 초기 이웃 반경, H_{inter} 는 변화되는 이웃 반경의 변화량이다.

표 1. KNN 학습 횟수에 따른 기호 영상 정규화
Table. 1 Normalization of Musical Symbols w.r.t. KNN Learning Cycles











기호 영상	정규화된 영상	학습 횟수
		500
		1,000
		2,000
		5,000
		10,000

본 논문에서는 추출된 기호 영상에서 픽셀의 값이 0인 각 점의 x,y 좌표를 입력 벡터로 사용한다. 좌표값은 [0,1]사이의 실수값으로 정규화 하며 KNN의 학습을 통하여 안정화된 가중치를 64×64 의 영상에 매칭시켜 정

규화된 영상을 추출한다. 일반적으로 KNN의 학습 과정은 매우 큰 비용을 요구하는 단점이 있다[5]. 표 1에서 알 수 있듯이 학습 횟수에 따라 정규화된 이미지가 안정화되어 가는 것을 확인할 수 있다. 하지만 입력 영상의 크기나 형태에 따라 안정화되기 위한 학습 횟수가 달라진다. 이러한 문제는 KNN이 가지는 자기 조직화 과정에서 발생하는 것으로 적절한 학습 횟수를 정의하기가 애매하다[5].

표 1의 2000회 이상에서 정규화된 영상과 비교하면 표 2의 영상이 많은 차이를 보이는 것을 알 수 있다. 하지만 동일한 기호 영상에 대하여 반복적으로 학습한 결과, 표 3과 같이 일정하게 정규화된 영상을 얻을 수 있었다.

표 2. $\sum\Delta$ 이 0.0001이하인 경우의 정규화
Table. 2 Normalization with $\sum\Delta < 0.0001$

기호 영상	정규화된 영상
	
	
	
	
	

ART-1은 자율 학습 신경망으로 0과 1의 이진 데이터 패턴을 군집화 하고 저장된 패턴을 즉시 회상할 수 있는 알고리즘이다. ART-1은 경계 변수를 조정하여 필요에 따라 활용할 수 있는 장점이 있다[7,8].

표 3. 동일한 기호 영상의 정규화
Table. 3 Normalization of same Symbols

기호 영상	정규화된 영상
	
	
	
	

하지만 기존의 ART-1은 식 (7)과 같은 유사성 검증 방법을 사용하는데 이 방법은 1의 값만 영향을 미치게 되어 패턴들을 명확히 분류할 수 없는 문제점이 발생한다. 식 (8)은 이러한 문제점을 해결하기 위해 Exclusive NOR의 Norm 비율을 사용하여 유사성 검증 방법을 개선한 ART-1의 유사성 검증 방법이다[6]. 본 논문에서는 개선된 ART-1을 적용하여 SOM에 의해 정규화된 패턴들을 학습하고 인식에 적용한다.

$$\frac{\|T \cdot X\|}{\|X\|} \tag{7}$$

$$\frac{\|T \odot X\|}{M} \tag{8}$$

식 (7)에서 T는 저장 패턴, X는 입력 패턴이다. 식 (8)에서 M은 입력 패턴의 크기이다.

III. 실험 및 결과 분석

본 논문에서는 단순히 레스터 정보에 대한 레스터 정보로의 정규화 실험이므로 벡터화 방법은 제외하였다. 본 실험은 기호를 학습하기 위한 전처리 과정의 특징을 살펴보는 부분으로 SOM에 의한 기호 정규화의 효율성

을 확인하기 위해 진행하였다. 정규화 후, 학습 과정은 모두 동일하게 개선된 ART-1을 적용하였으며, 경계 변수를 달리하여 경계 변수에 따른 클러스터의 생성 개수도 비교하였다. 학습에 사용된 패턴은 그림 3과 같고 각기 모양이 다른 다양한 종류의 기호들을 대상으로 생성된 클러스터의 비율을 비교 및 분석하였다.

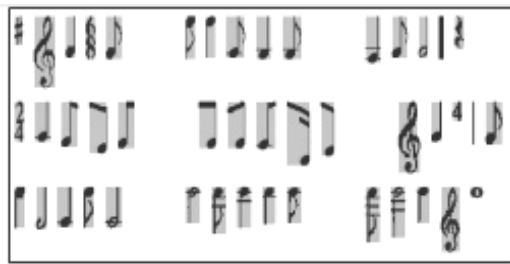


그림 3. 학습 패턴
Fig. 3 Training Patterns

기호의 정규화로 인한 원 정보의 소실은 결국 유사한 형태의 정규화로 인하여 생성되는 클러스터의 증가량을 줄어든다. 따라서 표 4에서 알 수 있듯이 SOM에 의한 정규화가 원 정보의 소실이 적다는 것을 확인할 수 있다.

표 4. 정규화에 따른 클러스터링
Table. 4 Clustering after Normalization process

경계 변수	적용 알고리즘	생성된 클러스터/패턴 수
0.85	단순 이미지 리사이징	39 / 45
	세선화 적용	39 / 45
	SOM 정규화	44 / 45
0.70	단순 이미지 리사이징	31 / 45
	세선화 적용	33 / 45
	SOM 정규화	42 / 45
0.55	단순 이미지 리사이징	24 / 45
	세선화 적용	29 / 45
	SOM 정규화	36 / 45

본 논문에서 제안하는 SOM과 개선된 ART-1을 적용하여 음표를 인식하는 방법의 인식률을 실험하였다. 제안된 방법을 검증하기 위해 본 논문에서 적용한 전처리 과정으로 추출된 기호들을 대상으로 학습을 수행하였

다. 그림 4는 인식이 완료된 악보이다. 그림 4에서 빨간색 사각형은 음을 나타내는 각 음표를 나타내며, 빨간색 사각형은 그 외의 기호들이다. 그리고 -1로 표시된 기호는 추출은 되었으나 인식이 실패한 기호이다. 즉, 제안된 방법으로 추출된 기호 중 하나의 기호를 제외하고 모든 기호들이 인식되었음을 그림 4를 통해 확인할 수 있다.



그림 4. 인식된 결과
Fig. 4 Recognition Result

IV. 결론

본 논문에서는 SOM과 개선된 ART-1을 적용하여 음표를 학습하고 인식 방법을 제안하였다.

악보 기호 추출을 위해 스캔된 악보 이미지에 호프 변환과 Otsu's 이진화를 적용하였고, 히스토그램 분석을 통해 악절을 구분하였다. 구분된 각각의 악절에 오선을 제거한 후, 레이블링을 수행하여 음표 성분을 분리하였다. 추출된 악보 기호들은 SOM 알고리즘을 적용하여 일정한 크기로 정규화 하였으며, 정규화된 악보 기호들은 개선된 ART-1 알고리즘에 적용하여 학습한 후에 인식에 적용하였다. 제안된 방법으로 실험한 결과, 제안된 방법이 음표 인식에 효율적인 것을 확인하였다.

향후 연구 과제로는 전처리 과정에서 추출되는 음표 성분이 정보 손실 없이 분명하게 추출되도록 개선할 것이며, 개선된 ART-1보다 인식률이 더 향상된 ART-2나 SVM 알고리즘을 적용하여 인식률을 개선할 것이다. 또한 SOM 알고리즘의 단점인 느린 속도로 인하여 전체적인 퍼포먼스가 느려지는 문제점이 있다. 본 논문에서는 $\Sigma\Delta$ 를 사용하여 어느 정도 속도를 향상시켰지만 실생활에 적용하기에는 인식 속도가 느리다. 따라서 SOM 알고리즘의 속도 향상을 위한 연구도 진행할 것이다. 더불어 최근 연구에서 사용되는 악보 규칙성 활용을 이용하면 인쇄 악보뿐 아니라 손으로 작성된 악보나 복잡한 음악 기호의 인식도 가능할 것이다.

참고문헌

- [1] J. Gomes, L. Velho, Image Processing for Computer Graphics, Springer-Verlag New York, Inc., 1997.
- [2] S. Lee, J. Shin, "Recognition of Music Scores using Neural Networks", *Journal of the Korean Institute of Information Scientists and Engineers*, Vol.21, No.7, pp.1358-1366, 1994
- [3] J. M. Yoo, G. H. Kim, G. S. Lee, "Music Recognition by Partial Template Matching," *Journal of Korea Contents Association*, Vol.8, No.11, pp.85-93, 2008.
- [4] K. B. Kim, Y. W. Woo, "Musical Score Recognition Using Hierarchical ART2 Algorithm." *Journal of The Korea Institute of Maritime Information & Communication Sciences*, Vol.13, No.3, pp.1997-2003, 2009.
- [5] J. H. Cho, H. J. Park, K. B. Kim, "Vector Quantization Using Enhanced SOM Algorithm," *Lecture Notes in Computer Science*, Vol. 3320, pp.172-175, 2005.
- [6] K. B. Kim, S. W. Jang, C. K. Kim, "Recognition of Car License Plate by Using Dynamical Thresholding Method and Enhanced Neural Networks", *Lecture Notes in Computer Science*, Vol. 2756, pp. 309-319, 2003.
- [7] B. Yu, A. K. Jain, "Lane boundary detection using a multiresolution Hough transform", *Proceedings of*

International Image Processing Conference, Vol.2, Iss.26-29, pp.748 -751, 1997.

- [8] M. Georipoulos, G. L. Heileman and J. Huang, "Properties of Learning Related to Pattern Diversity in ART1," *Neural Networks*, Vol.4, pp.751-757, 1991.

저자소개



김광백(Kwang-Baek Kim)

1999년 : 부산대학교
전자계산학과(이학박사)
1997년~현재 신라대학교
컴퓨터공학과 교수

현재 한국정보통신학회총무 부회장

2007년~현재 The Open Artificial Intelligence Journal
(USA) Editor

※관심분야: 퍼지 논리, 영상 처리, 유전자 알고리즘,
의료정보시스템, 생물정보학