JKIICE

Short Paper

한국정보통신학회논문지 Vol. 26, No. 7: 1098~1101, Jul. 2022

손사보 악보의 광학음악인식을 위한 CNN 기반의 보표 및 마디 인식

Staff-line and Measure Detection using a Convolutional Neural Network for Handwritten Optical Music Recognition

Jong-Won Park[1*] · Dong-Sam Kim[2] · Jun-Ho Kim[3]

[1*]CTO, R&D Center, Juice Inc., Seoul, 04521 Korea
[2]Principal Researcher, R&D Center, Juice Inc., Seoul, 04521 Korea
[3]CEO, Juice Inc., Seoul, 04521 Korea

ABSTRACT

With the development of computer music notation programs, when drawing sheet music, it is often drawn using a computer. However, there are still many use of hand-written notations for educational purposes or to quickly draw sheet music such as listening and dictating. In previous studies, OMR focused on recognizing the printed music sheet made by music notation program. the result of handwritten OMR with camera is poor because different people have different writing methods, and lens distortion. In this study, as a pre-processing process for recognizing handwritten music sheet, we propose a method for recognizing a staff using linear regression and a method for recognizing a bar using CNN. F_1 scores of staff recognition and barline detection are 99.09% and 95.48%, respectively. This methodologies are expected to contribute to improving the accuracy of handwriting.

Keywords : Convolutional Neural Network, Handwritten music sheet, Measure detection, Staff-line detection

Ⅰ. 서 론

OMR(Optical Music Recognition)은 인쇄되거나 손으로 쓴 악보를 스캐너나 카메라를 통해 수집한 데이터를 악보의 구조와 음악 개체를 인식하여, 음악 내용을 이해할 수 있도록 하는 기술이다[1]. 컴퓨터 사보 프로그램의 발달로 인하여 악보를 그릴 때, 컴퓨터를 이용하여 그리는 경우가 많다. 하지만, 아직도 교육적인 목적이나 듣고 받아적는 등의 경우에 빠르게 악보를 그리기 위해서는 손을 이용한 사보를 하고 있다. 기존에 OMR (Optical Music Recognition)은 컴퓨터 사보프로그램을 이용하여 출력된 결과물을 스캐너를 이용하여 인식하는 것에 초점이 맞추어져 있다. 손으로 작성된 악보를 사진 촬영하는 경우 화질 및 렌즈의 왜곡으로 인하여 인식 결과가 좋지 못하고, 사보에 사용된 필기구에 의해서도 결과물의 차이가 크다.

본 연구에서는 카메라를 통해서 사용자가 직접 촬영한 이미지를 이용하여 손사보 인식을 수행하려고 한다. 이중 손으로 사보한 악보를 인식하기 위한 전처리 과정으로 선형 회귀와 선을 구분하는 알고리즘을 이용한 보표를 인식하는 방법과 CNN(Convolutional Neural Network)을 이용하여 마디줄을 인식하고, 마디를 구분하는 방법을 제안한다. 이 방법론은 카메라를 이용하여 촬영된 손사보의 정확도를 높이는데 기여할 수 있을 것으로 기대한다.

Ⅱ. 손사보 광학음악 인식

손사보 광학 인식은 딥러닝의 출현으로 인해 다양한 접근 방법이 제안되었다. F. Alirezazadeh과 M. R. Ahmadzadeh은 손사보에서 보표를 찾고 제거하기 위한 연구를 수행하였지만, 인공지능을 이용한 것이 아닌 단순 휴리스틱(heuristic)을 이용하였다[2]. A. J. Gallego et al. 과 F. J. Castellanos et al.은 머신러닝 알고리즘 중 하나인 Auto-encoder를 사용하여 손사보의 보표를 제거하는 연구를 수행하였다[3, 4].

손사보를 위한 데이터 셋으로 MUSCIMA++ [5],

Received 13 June 2022, **Revised** 17 June 2022, **Accepted** 28 June 2022

*** Corresponding Author** Jong-Won Park(E-mail:jason.park@juice.co.kr, Tel:+82-70-7700-0200)
CTO, R&D Center, Juice Inc., Seoul, 04521 Korea

Open Access http://doi.org/10.6109/jkiice.2022.26.7.1098 print ISSN: 2234-4772 online ISSN: 2288-4165

Baró Single Stave Dataset[6] 등이 있다. 하지만 MUSCIMA++과 Baró Single Stave Dataset의 경우, 고해상도이고, 렌즈의 왜곡이 없으며, 필기구의 종류가 일정하여, 카메라에서 직접 사진을 찍어서 OMR를 수행하는 것을 위한 학습용 데이터로 사용하기에 데이터의 성질이 다르다. 때문에 본 논문에서는 355장의 악보를 휴대폰 카메라를 포함한 다양한 카메라로 불규칙하게 촬영하고, 이를 태깅하여 사용하였다.

OMR을 위한 다양한 처리 방법론과 절차가 제시되어 있지만, 본 연구에서는 보표와 마디를 인식하고, 그 후 마디 별로 음악 객체를 인식한다. 마디별로 인식된 음악 객체를 병합하고, 마디별로 병합된 데이터를 통합하고, 마지막으로 음악 정형화를 위한 표준 포맷인 musicxml로 변환한다. 그림 1은 본 논문에서 제시하는 OMR 처리 과정이다.

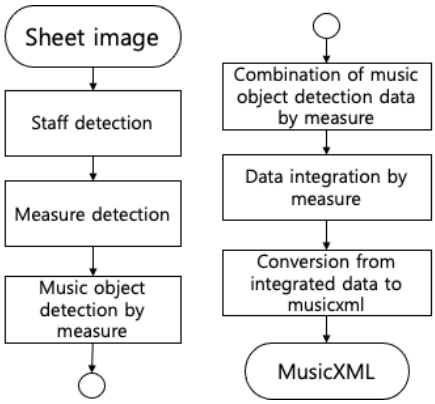


Fig. 1 Process steps of optical music recognition

Ⅲ. 보표인식

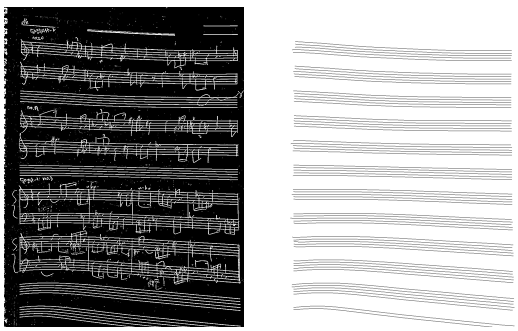
3.1. 보표인식 알고리즘

보표인식을 위해서 먼저 이미지를 입력으로 받아 흑백이미지로 변환한다. 흑백이미지로 변환하기 위해서 NxN 픽셀 마다 그레이 스케일(gray scale) 값이 전체 이미지 평균의 95%보다 크면 흰색으로 아니면, 검정색으로 처리하여 그림자를 제거한다. 개발시 N의 값은 10으로 사용하였다. 선을 예측하기 위한 모델로서, 3개의 선형회귀를 직렬연결하여 이용하였다. 첫 번째 선형회귀의 입력은 이미지 사이즈 즉, 10x10이고 출력은 64이다. 두 번째 선형회귀 입력은 64, 출력은 32이고, 마지막 모델의 입력은 32, 출력은 4이다. 좌측 20% 영역에 흰색점을 개수만큼 선 객체를 생성하고, 그림2에서와 같이 각 점에서 우측 값은 (a)흰색, (b) 검정색, (c) 끝 이렇게 3가지 경우를 갖는다. 만약, 흰색인 경우 즉시 확장하고, 검정색인 경우 예측하여 우상, 우, 우하 중 한방향으로 확장 한다. 녹색은 확장된 결과를 보여준다. 예측결과가 end인 경우 라인 확장을 종료한다.

(a) black (b) white (c) end

Fig. 2 3 Cases of staff prediction

이미지의 가로 길이의 50%이상인 선만을 필터링하고, 이미 발견한 선과 확장하고 있는 선이 중복될 경우에는 끝 지점에서 길이가 가장 긴 선만을 남긴다. 마지막으로 인식된 라인을 5개씩 묶어서 보표로 처리한다. 그림 3-(a)는 보표 인식을 위한 전처리 결과이고, 그림 3-(b)는 보표 인식 결과를 나타낸다.



(a) (b)

Fig. 3 (a) Preprocessing result, (b) Staff detection result

3.2. 실험결과

구현언어는 python, 머신러닝 프레임워크는 pytorch를 이용하였다. 보표인식을 위해 학습(learning)에 사용한 데이터 중 보표는 총 1472개이고 확인(validation)에 사용한 데이터는 164개이다. 실험결과는 보표인식과 마디 인식의 성능을 측정지표로 사용되는 F1를 이용하였다[7]. 수식1에서 TP는 True Positive, FP는 False Positive, FN은 False Negative를 나타낸다.

$$F_1 = \frac{2 \bullet TP}{2 \bullet TP + FP + FN} \quad (1)$$

테스트 데이터로는 1117개의 오선보를 이용하였고, TP는 1097, FN은 20, FP는 0였고, F_1값은 99.09%로 계산된다.

Ⅳ. 마디인식

4.1. 마디인식 알고리즘

앞에서 인식한 보표에 속한 선의 모든 점에 대해서 마디줄인지를 예측하기 위해서, 그림 4와 같이 2개의 컨볼루션 계층(Convolution layers)과 그림 5와 같이 3개의 완전 연결 계층(Fully-connected layers)으로 컨볼루션 신경망(Convolutional neural network)을 구성하였다. 컨볼루션 신경망의 입력은 101x101픽셀의 이미지가 입력으로 들어가며, 첫 번째 컨볼루션 계층에서는 5x5 컨볼루션 필터 6채널과 Relu 활성함수(activation function)에 이어서, 2간격(2 stride) 2x2 최대 풀링(Max pooling)을 사용하였다. 두 번째 컨볼루션 계층에서는 5x5컨볼루션 필터 16채널과 Relu 활성함수에 이어서 2간격(2 stride) 2x2 최대 풀링(Max pooling)을 사용하였다.

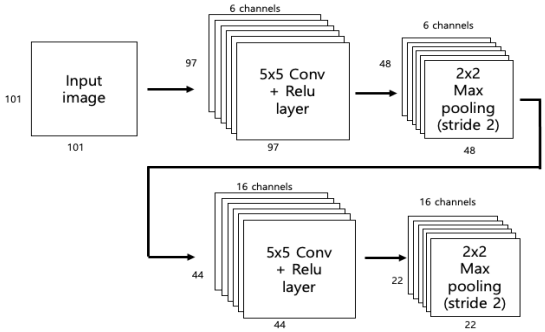


Fig. 4 Convolution layers

완전 연결 계층은 그림 5와 같이 구성되어 있다. 완전 연결 계층의 입력으로 사용하기 위해서, 컨볼루션 계층의 결과 구조인 16x22x22를 flatten하여 1차원인 7744개의 노드로 변환한다. 완전 연결 계층에서는 120노드, 84노드를 중간 결과물로 이용하고, Relu 활성함수를 이용하였다. 최종적으로 출력값으로 0과 1만을 갖도록 softmax 함수를 이용하였고, 이때 비용함수(cost function)를 최소화 하기 위한 최적화 알고리즘으로는 Adam 최적화 함수(Optimizer)를 사용하였다.

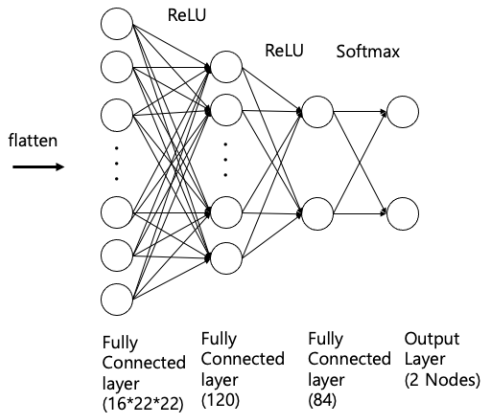


Fig. 5 Fully-connected layers

예측을 위해서 이미지를 입력하면, 보표 인식 단계에서 인식한 선의 점들을 마디 인식의 네트워크에 입력으로 사용하였다. 이 네트워크의 출력값은 0또는 1이다. 1이면, 그림6의 파란 부분과 같이 마디줄에 해당하는 점으로 판별하고, 0이면, 무시한다. 그림 7과 같이 보표 각 선에서 1로 인식된 점을 묶고 이어서 마디줄을 생성한다. 마디줄을 기준으로 하여 마디를 구분한다.

Fig. 6 Prediction of barline point

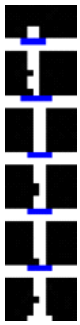
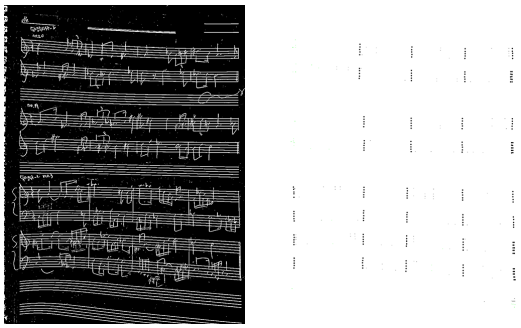


Fig. 7 merging of barline points

그림 8-(a)는 입력으로 사용한 이미지이고, 그림 8-(b)는 마디줄 인식 결과를 보여준다.



(a) (b)

Fig. 8 (a) Input image of barline detection, (b) Barline detection result

4.2. 실험결과

마디 인식 또한, 구현언어는 python, 머신러닝 프레임워크는 pytorch를 이용하였다. 마디인식을 위해 학습(learning)에 사용한 마디줄 데이터는 총 4329개이고 확인(validation)에 사용한 데이터는 482개이다. 테스트 용 데이터는 2729개를 이용하였다. 실험결과의 성능 측정 지표는 보표인식과 같이 F_1를 이용하였다. 측정결과 TP는 2631개, FN은 98개, FP는 151개였으며, 최종 F_1 수치는 95.48%로 보였다.

V. 결 론

본 논문에서는 손사보 인식을 위한 전처리 단계인 보표 인식과 마디 인식 방법에 대해서 제시하였다. 보표 인식은 선형 회귀 방법과 선을 선별하는 알고리즘을 사용하였다. 마디 인식은 CNN을 이용하여, 마디줄을 검출하고, 이를 기반으로 마디를 구분하는 알고리즘을 사용하였다. 보표 인식과 마디 인식의 성능을 측정하기 위한 지표로서 F_1 점수를 이용하였다. 보표 인식과 마디 인식의 F_1 점수는 각각 99.09%과 95.48%를 나타냈다. 본 연구의 후속 연구로서, 음악 개체 감지, 감지된 음악 개체를 재구성하는 방법에 대해서 수행할 계획이다.

ACKNOWLEDGEMENT

This research is supported by Ministry of Cultures, Sports and Tourism and Korea Creative Content Agency(Project Number: R2021050006)

REFERENCES

[1] A. Pacha, K. -Y. Choi, B. Coüasnon, Y. Ricquebourg, R. Zanibbi, and H. Eidenberger, "Handwritten Music Object Detection: Open Issues and Baseline Results," in *Proceeding of 13th IAPR International Workshop on Document Analysis Systems*, Vienna, Austria, pp. 163-168, 2018. DOI: 10.1109/DAS.2018.51.

[2] F. Alirezazadeh and M. R. Ahmadzadeh, "Effective staff line detection, restoration and removal approachfor different quality of scanned handwritten music sheets," *Journal of Advanced Computer Science & Technology*, vol. 3, no. 2, pp. 136-142, Jun. 2014. DOI: 10.14419/jacst.v3i2.3196.

[3] A. J. Gallego and J. Callvo-Zaragoza "Staff-line Removal with Selectional Auto-Encoders," *Expert Systems with Applications*, vol. 89, pp. 138-148, Dec. 2017. DOI: https://doi.org/10.1016/j.eswa.2017.07.002.

[4] F. J. Castellanos, J. Calvo-Zaragoza, G. Vigliensoni, and I. Fujinaga, "Document Analysis of Music Score Images with Selectional Auto-encoders," in *Proceeding of the 19th International Society for Music Information Retrieval Conference*, Paris, France, pp. 256-263, 2018. DOI: 10.5281/zenodo.1492397.

[5] A. Fornés, A. Dutta, A. Gordo, and J. Lladós, "CVC-MUSCIMA: A Ground-truth of Handwritten Music Score Images for Writer Identification and Staff Removal," *International Journal on Document Analysis and Recognition*, vol. 15, no. 3, pp. 243-251, Jun. 2012. DOI: 10.1007/s10032-011-0168-2.

[6] A. Baró, P. Riba, J. Calvo-Zaragoza, and A. Fornés. "From Optical Music Recognition to Handwritten Music Recognition: a Baseline," *Pattern Recognition Letters*, vol. 123, pp. 1-8, May. 2019. DOI: 10.1016/j.patrec.2019.02.029.

[7] J. Calvo-Zaragoza, A. Pertusa, and J. Oncina, "Staff-line detection and removal using a convolutional neural network," *Machine Vision and Applications*, vol. 28, pp. 665-674, May. 2017. DOI: 10.1007/s00138-017-0844-4.