

비전 기반 피아노 자동 채보 시스템

Vision-Based Piano Music Transcription System

박 상 옥*, 박 시 현*, 박 천 수*

Sang-Uk Park*, Si-Hyun Park*, Chun-Su Park*

Abstract

Most of music-transcription systems that have been commercialized operate based on audio information. However, these conventional systems have disadvantages of environmental dependency, equipment dependency, and time latency. This paper studied a vision-based music-transcription system that utilizes video information rather than audio information, which is a traditional method of music-transcription programs. Computer vision technology is widely used as a field for analyzing and applying information from equipment such as cameras. In this paper, we created a program to generate MIDI file which is electronic music notes by using smart-phone cameras to record the play of piano.

요 약

현재 상용화된 악보 채보 프로그램은 오디오 정보를 기반으로 채보를 진행한다. 이러한 기존 채보 프로그램은 환경 의존성, 장비 의존성, 시간 지연이라는 단점을 지니고 있다. 본 논문은 기존의 오디오를 이용하여 채보를 방식을 지양하고, 연주 영상을 분석하여 채보를 진행하는 컴퓨터 비전 기반 악보 채보 시스템을 제안한다. 제안하는 악보 채보 시스템은 대중화된 스마트폰 카메라를 활용하여 피아노 연주를 촬영하고, 이를 분석하여 자동으로 전자 악보인 미디파일을 생성하는 방식으로 동작한다. 컴퓨터 실험에서 제안하는 악보 채보 시스템은 95.6%의 정확도로 연주된 음계를 채보하는 것으로 조사되었다.

Key words : Computer Vision, SVM, Music-transcription System, Smartphone Camera, Object Detection

1. 서론

현대의 가장 대중적이며 보편적인 문화로 음악이 자리매김하고 있다. 대중은 단순한 음악의 소비를 넘어서 창작에 대한 욕구를 보이고 있으며, 이러한 대중들의 욕구에 맞춰 음악 창작을 보조하는 프로그램들이 다양하게 사용되고 있다. 대부분의 음악 창작 보조 프로그램은 다음과 같은 두 가지 형태를 가진다. 하나는 실물 악기를 대체하여 사용하는 소프트웨어 악기의 형태이고 다른 하나는 악기의 연

주를 전자 악보인 미디파일의 형태로 변환해주는 형태이다.

악기의 연주를 미디파일로 변환해주는 프로그램을 악보 채보 프로그램이라 칭한다. 전통적인 악보 채보 방식은 전문가가 음악을 직접 듣고 보조 도구를 사용하여 수동으로 채보하는 방식으로 진행되어 왔다. 이러한 단점을 보완하기 위해 음계와 박자와 같은 음악적 특징을 자동으로 인식하여 악보로 만들어주는 오디오 기반의 자동 채보 프로그램이 개발되어 왔다. 본 연구에서는 한발 더 나아가

* Dept. of Computer Education, SungKyunKwan University

★ Corresponding author

E-mail : cspk@skku.edu, Tel : +82-2-760-0678

※ This paper was supported by Sungkyun Research Fund, Sungkyunkwan University, 2018

Manuscript received Mar. 20, 2019; revised Mar. 25, 2019; accepted Mar. 25, 2019.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

오디오 기반이 아닌 영상 기반 즉 컴퓨터 비전을 기반의 악보 자동 채보 시스템(Vision-Based Music Transcription System)을 개발하였다.

II. 기존 연구 동향

현재까지 상용화된 모든 자동 채보 프로그램은 음향 정보를 처리하여 악보를 생성하는 오디오 기반의 형태이다. 이러한 경우, 주변 환경에 소음이 존재할 때, 생성되는 악보가 부정확해지는 문제점이 있으며 고가의 미디 장비가 있어야 채보가 가능하다는 단점을 지니고 있다. 또한 컴퓨터 비전을 활용한 연구들은 대부분 가상의 피아노를 연주하는 환경을 구축하는 방식[1], 혹은 Depth Camera를 이용하여 연주자의 제스처를 인식하는 방식[2, 3, 4]으로 진행되어 왔다.

1. Chordify

Chordify는 Youtube 등과 같은 동영상 스트리밍 사이트로부터 오디오 정보를 추출 및 재구성하여 코드 악보를 생성해주는 프로그램이다. Chordify는 멜로디의 세부 계이름은 알 수 없으며 전반적인 코드 진행만을 알 수 있다.

2. Audioscore

Audioscore는 녹음된 음악의 음향 정보를 처리하여 악보를 만들어주는 프로그램이다. 해당 프로그램은 3분 분량의 오디오 파일을 채보하기 위해서 15분 정도의 시간이 소요되며, 음량 확보와 음색의 조정 정도를 뜻하는 마스터링 상태에 따라 정확도의 차이가 발생한다.

III. 제안 방법

Vision-Based Music Transcription(VBMT) 시스템은 영상을 기반으로 피아노 연주를 감지하여 이에 해당하는 악보를 자동으로 생성한다. VBMT 시스템은 크게 두 가지 모듈로 구성되어 있으며 입력으로 피아노 연주 영상을, 출력으로는 전자 악보 형태인 미디파일(MIDI file)을 생성한다. 그림 1은 VBMT 시스템의 pipeline을 구조화한 블록도이다.

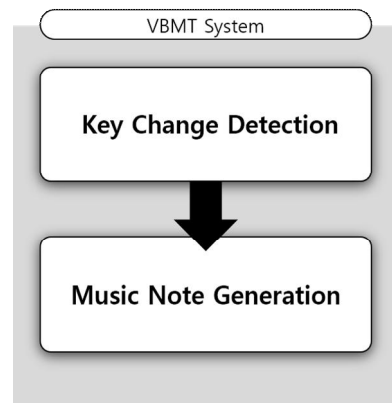


Fig. 1. Main pipeline of the VBMT System.
그림 1. VBMT 시스템의 주요 파이프라인

1. Key Change Detection

해당 모듈은 피아노 건반의 타건을 감지하는 동작을 담당한다. 세부 모듈로는 키보드 탐색(Keyboard Detection, KD), 건반 매핑(Key-Note Mapping, KNM), 타건 판정(Key Press Detection, KDP), 기준 이미지 생성(Background Image Generation, BIG), 손 탐지(Hand Detection, HD)의 5가지 세부 모듈로 구성되어 있다. 그림 2는 KCD 모듈의 세부 구성을 나타낸 블록도이다.

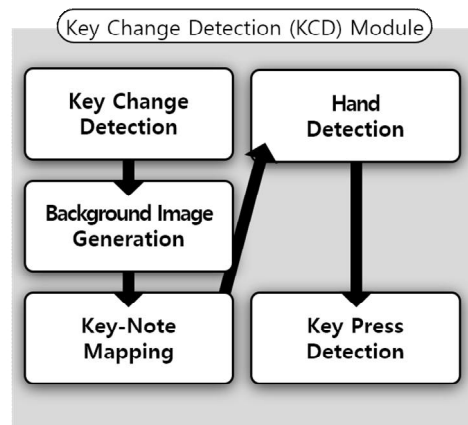


Fig. 2. Pipeline of the Key Change Detection Module.
그림 2. KCD 모듈의 주요 파이프라인

가. 키보드 탐색

키보드 탐색은 우선 입력 영상으로부터 기준이 되는 배경영상(Background-Image)을 추출하고 SVM(Support Vector Machine)을 통해 입력 영상에서의 키보드 위치를 탐색한다. 탐지된 키보드 영역은 관심영역(ROI)화 하여 다른 모듈로 전달된다[5].

나. 건반 매핑

건반 매핑은 총 4가지 단계를 거친다. 우선 어댑티브 스레시홀딩(Adaptive Thresholding, 가변 이진화) 연산을 한다. 빛의 반사 여부를 최소화 하고, 비슷한 영역을 하나의 영역으로 포함시켜 건반의 윤곽을 검출한다. 그 후 캐니 에지(Canny Edge) 검출을 통해 픽셀 단위로 외곽을 표현한다. 캐니 에지를 통해 산출된 결과물로 모폴로지 연산(Morphological Operation)을 진행하여 각각 건반에 대한 독립성을 확보하고 건반의 영역을 구분한다. 건반을 탐지할 때에는 흑건을 먼저 탐지한다. 백건의 경우 경계가 모호하여 탐지가 힘들지만, 흑건의 경우 백건 사이에 있기 때문에 그 경계가 분명하여 탐지가 수월하기 때문이다. 그림 3을 보면 원본 키보드 입력 영상과 흑건만 선택지를 통하여 백건만의 영역들이 확연하게 구분된 것을 확인할 수 있다.

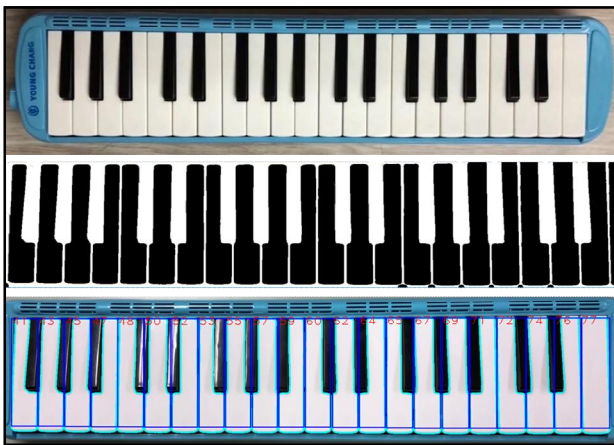


Fig. 3. Stages of the detecting keys by using black-key detection.

그림 3. 흑건 탐지를 통한 백건 검출 단계

다. Key Class 인스턴스 생성

건반 매핑 이후 컨투어(Contour) 연산을 통해 백건만의 영역을 독립적으로 구분 짓는다. 컨투어 연산을 통해 구분지은 각각의 영역들은 키보드의 높이와 비교했을 때 2/3 이하의 높이 값을 갖는 영역들을 제거함으로써 흰색 건반에 대한 정보만 추출할 수 있다. 각각의 건반에 대한 영역을 확정지은 후 타건 시 발생하는 키보드의 그림자를 탐지하기 위해 마스크(Mask)를 적용한다. 마스크는 타건 시 발생하는 키보드 전체의 진동으로 인한 노이즈와 손의 움직임으로 인해 발생하는 노이즈를 제거하기 위해 사용된다.

라. YCrCb Hand Detection

타건 판정 시 기준 영상과 직전 영상과의 차분 픽셀 변화를 이용하기 때문에 손을 정확하게 추적해서 손의 움직임을 타건 판정에서 제거하는 것은 필수적이다. YCrCb 색상 값을 활용하여 손의 색상 범위를 입력한다. YCrCb는 적은 용량으로 많은 범위의 색상이 표현되기 때문에 손 검출에 용이하다 [6, 7].

마. Key Press Detection

Key Press Detection은 입력 영상에서 건반이 눌렸는지 다시 떨어졌는지를 판단하는 모듈이다. 타건이 판정되었을 경우, 해당 정보를 벡터 값으로 저장한다.

2. Music-Note Generation

오픈소스인 Midifile을 사용하여 최종 음계 정보가 들어있는 벡터 값을 전자 악보인 MIDI의 형태로 변환한다[8]. 벡터에 담긴 정보를 변환하기 전에 타건 판정 결과 벡터를 후가공한다. 박자에 대한 임의의 값을 설정해 박자를 정규화하여 일괄된 박자의 형태를 갖춘 악보를 생성한다.

IV. 결론

본 연구에서는 컴퓨터 비전 기술을 활용하여 피아노 연주를 자동으로 악보로 만들어주는 비전 기반 악보 채보 프로그램을 제작하였다. 기존의 오디오 기반의 악보 채보 프로그램은 조율이 되지 않은 악기로 연주할 때 결과 값이 매우 부정확해진다. 단점과 지연시간이 매우 높다(3분 음악을 채보하기 위해 15분 소요)는 단점을 갖고 있었다.

본 연구에서 개발한 비전을 기반으로 하는 악보 채보 시스템인 VBMT 시스템은 지연시간이 낮은 장점을 갖고 있다. 타건과 동시에 타건 판정이 이루어지기 때문에 피아노 연주를 마치고 바로 결과를 확인할 수 있다. 또한, 악기의 조율 여부와는 상관없이 연주자의 연주만을 통해 타건 판정을 진행하기 때문에 오디오 기반의 악보 채보 프로그램보다 정확도를 향상시킬 수 있다.

표 1을 보면, 동요 '비행기(Mary had a little lamb)' 노래를 조도가 좋은 환경에서 키보드로 연주했을 경우, 대부분의 음계가 정확하게 악보로 표시되었다.

또한 양손 모두가 사용되며 동시에 여러 화음이 들어가는 복잡한 형태를 지닌 ‘첫가락 행진곡(Celebrated Chop Waltz)’을 연주하였을 경우에도 대부분의 음계가 정확하게 악보로 산출되었다.

박자의 경우, 연주자가 정확한 박자를 기계적으로 연주할 수 없음을 고려하여 전체 악보에서 해당 음이 차지하는 비율을 계산하여 일치율을 산출하였다. 하지만 동일한 음이 연속적으로 이어지는 경우에 해당 음계의 박자가 짧아지는 현상이 발생하여 박자 정확도를 떨어트렸다.

Table 1. Correspondent rate of the VBMT System.

표 1. VBMT 시스템의 악보 일치율

Song (Input Video)	score(%)	rhythm(%)
Marry Had a Little Lamb	98.0	32.3
Celebrated Chop Waltz	93.2	35.4

또한 정확도를 높이기 위해서는 충분한 조도 환경이 조성되어야 한다는 단점을 지니고 있다. 조도 환경이 좋지 않는 대부분의 경우에 키보드가 탐지되지 않으며, 타건 판전이 불가능하다. 뿐만 아니라 피아노 연주의 범위가 모두 촬영 영역에 포함되어야 하며 TOP view로 촬영되어야 한다는 단점을 지니고 있다. 향후 다양한 조도환경에서의 실험이 필요하며 카메라 제원에 따른 비교 연구가 수반되어야 본 연구에서 제시한 시스템의 질적 향상을 기대할 수 있다.

References

[1] Broersen and Nijholt, “Developing a Virtual Piano Playing Environment,” *IEEE International conference*, 2002.

[2] Q Yang and G Essl, “Augmented Paino performance using a Depth Camera,” *NIME*, 2012.

[3] Aristotelis Handjakos, “Pianist Motion Capture with the Kinect Depth Camera,” 2012.

[4] A. Oka and M. Hashimoto, “Marker-Less Piano Fingering Recognition suing Sequential Depth Images,” *The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp. 1-4, 2013. DOI: 10.1109/FCV.2013.6485449

[5] R. Bhuvaneswari and Ravi Subban, “Novel object detection and recognition system based on points of interest selection and SVM classification,” *ScienceDirect, Cognitive System Research* 52, pp. 985-994, 2018. DOI: 10.1016/j.cogsys.2018.09.022

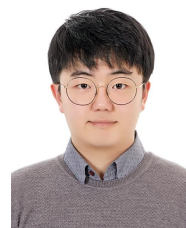
[6] G. Mano, Y. Wu, M. Hor and C. Tang, “Real-Time Hand Detection and Tracking against Complex Background,” *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 905-908, 2009. DOI: 10.1109/IIH-MSP.2009.133

[7] Hemlata Chavan and Prateek Gupta, “A Review on Hand Gesture Detection Using Combine HSI, YCbCr and Morphological Method Recognition,” *International Research Journal of Engineering and Technology*, vol. 03, no. 05, 2016.

[8] Craig Stuart Sapp, “C++ library for parsing Standard MIDI Files,” <https://midifile.sapp.org>.

BIOGRAPHY

Sang-Uk Park (Member)



2012 : graduate from Seil High School.
 2019 : attending 4th year in Dept. of Computer Education, SungKyunKwan University.
 2018 : Prime Minister Award, Global SW Contest.
 2019 : Excellence Award, SKKU Co-Deep-Learning.

Si-Hyun Park (Member)



2015 : graduate from Goyang Global High School.
 2019 : attending 4th year in Dept. of Computer Education, SungKyunKwan University.
 2018 : Prime Minister Award, Global SW Contest.
 2019 : Excellence Award, SKKU Co-Deep-Learning.

Chun-Su Park (Member)

2003 : BS degree in Electrical Engineering, Korea University.

2009 : PhD degree in Electronics and Computer Engineering, Korea University.

2017~Present : Professor in Dept. of Computer Education, Sungkyunkwan University.