

컴퓨터 비전 시스템에 의한 印刷樂譜의 認識과 演奏

(The Recognition of Printed Music Score and Performance Using Computer Vision System)

李明雨*, 崔宗秀*

(Myung Woo Lee and Jong Soo Choi)

要約

本論文에서는 印刷樂譜 畫像을 CCTV 카메라로써 마이크로 컴퓨터에 입력시켜, 이 畫像을 認識, 스피커로 노래를 내어주는 컴퓨터 비전 시스템에 관해 논하고 있다. 이때 樂譜畫像의 特徵抽出 및 認識에는 加算投影法이 적용되고, 그 對象 認識 範圍는 樂譜의 여러 요소 중에서 五線, 가디, 音標로 하고 있다. 아울러 實際 樂譜畫像을 취급할 때 반드시 고려되어야 할 前處理 및 雜音 除去 過程을 보였고, 認識된 音標로 和音을 내어주는 간단한 하드웨어 시스템을 구성했다. 그 結果, 양호한 認識率로 연주 가능성을 보였다.

Abstract

In this paper, a computer vision system, which catches printed music score image using CCTV camera and microcomputer, and then recognizes the image and performs the music with speaker, is discussed. Integral projection method is adopted for feature detection and recognition of the music score image. The range of recognition is confined to staves, perpendicular lines and musical notes including chord notes among the various kinds of elements of music score. The practical recognition algorithm considering noises, the preprocessing processes getting rid of noises are also showed, and simple hardware system playing chord is made. In the results, good recognition ratio and performance are obtained.

I. 序 論

컴퓨터에 視覺裝置를 부착하여, 이를 통해 외부의 畫像情報를 컴퓨터에 입력시키고 처리하는 디지털 畫像處理 분야¹⁾는 인간이 외부 세계의 많은 자극과 知識을 눈이라는 시각기관을 통해 얻고 뇌로 판단, 처리한다는 사실에 입각한 것이다.

일반적으로 畫像은 여러가지 다른 성질의 對象(pat-

tern)으로 구성되어 있으며, 이들을全體로써 이해하려는 人工知能 개념하의 畫像理解 (image understanding)에 관한 研究가 盛行되고 있다.²⁾ 또한 그에 앞서 人力畫像을 균일한 성질을 띤 對象에 한정시켜 제한된 대상만을 認識하는 패턴認識의 연구도 활발하다.^{3) 4)}

이러한 추세에 따라, 최근의 人工知能 개념을 도입하여, 컴퓨터에 視覺裝置를 부착, 주어진 樂譜畫像을 認識하고, 認識된 結果를 로봇 팔에 전달하여 無人自動演奏를 실현하기 위한 研究에 족히 관심을 갖을 수 있다. 그러나 그동안의 이 컴퓨터 뮤직 분야는 音樂情報를 코우드화한다든지, 또는 音樂 데이터 베이스 시스템을 구성하여 音樂情報를 分類 및 檢索하는

*正會員, 中央大學校 電子工學科
(Dept. of Electron. Eng., Chung-Ang Univ.)
接受日字: 1984年 11月 22日

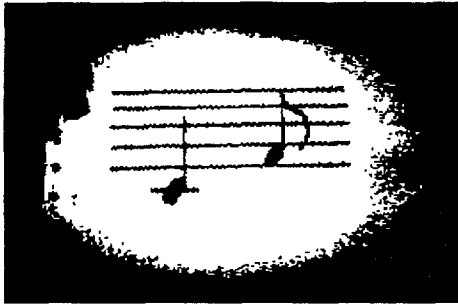


그림 4. 前處理 및 獨立孤立點 雜音を 除去한 뒤의 入力 畫像

Fig. 4. Input image after preprocessing and spot elimination.

그림 4는 上述의 前處理를 거친 뒤의 畫像이다. 여기서 樂譜信號 周圍의 검은 部位는 본 시스템에서 생기는 잡음이다. 이의 除去過程은 後述될 것이다.

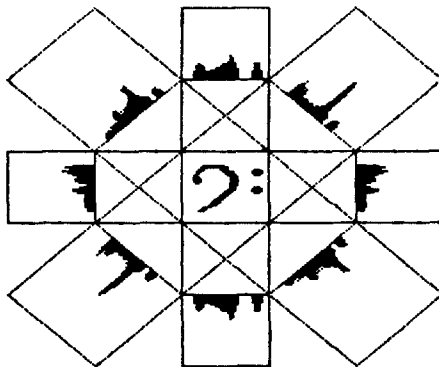


그림 5. 여러 각도의 加算技影
Fig. 5. Integral projections from various angles.

2. 各 五線들의 좌표 抽出

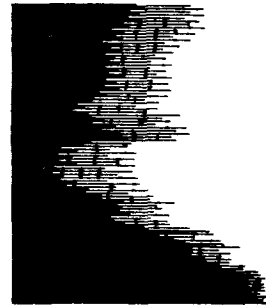
本 論文에서 認識 데이터를 얻는 過程에 加算投影法 (integral projection method)^[5]을 사용하고 있다. 이는 그림 5에서 처럼 가운데에 畫像信號가 있을 때, 임의의 방향에서(예를 들면, 0°, 45°, 90°, …등) 投影하여 그 投影線上에서 各 畫素가 갖고 있는 값을 모두 합하는 것이다.

그림 6에 2개의 2值 畫像에 대한 가로 방향 加算投影 例를 보이고 있다. 일반적인 얼굴 화상의 특징추출의 경우^[5]와는 달리, 樂譜 畫像은 가로 방향과 세로 방

향만으로 本 論文에서 의도하는 目的을 달성할 수 있다.



(a) Human face



(b) The horizontal projection of (a)



(c) Music score



(d) The horizontal projection of (c)

그림 6. 두 畫像의 가로 방향 加算技影
Fig. 6. The horizontal projections of two images.

五線 좌표 抽出의 경우는 가로 방향의 加算投影에 적당한 문턱치 처리를 하면, 五線의 좌표를 추출할 수 있다. 樂譜는 반드시 5 개의 가로선이 등간격으로 명확하게 그어져 있으므로, 그림 6(d)에서도 알 수 있듯이, 가장 큰 投影值(五線上에 따라 取해진 加算投影值)를 100으로 봤을 때 문턱치는 50이상 80이하 정도면 족하다. 여기서는 70으로 하고 있다.

그러나 人力畫像에 잡음이 많이 있는 경우는 상기 방법만으로는 어려가 발생한다. 예를 들어, 그림 4와 같이 樂譜信號 주위에 검게 싸고 있는 많은 잡음을 제거하기 위해서는, 먼저 수직방향의 加算投影을 얻은 다음, 그 投影值가 어느 임계값보다 크면 그 선상에서 모든 데이터를 소거한다. 왜냐하면, 그림 4에서 알 수 있듯이, 樂譜信號 주위의 검은 부위 投影值는 樂譜信號가 存在하는 부위에 비해 상대적으로 크기 때문이며, 따라서 어렵지 않게 그 임계값을 설정하여 불필요한 주위의 데이터를 없앨 수 있다. 이때 임계값의 설정은 배경잡음의 정도에 따라 정한다. 즉 잡음이 많을 때는 크게 그리고 적을 때는 작게 설정된다. 본 연구에서는 한 畫像의 全畫素數(237×192=45504) 對 人力畫像의 總畫素數(그림 4의 경우, 25318)의 比를 근거로 경험적으로 정하고 있다. 이렇게 하여 얻은 화상을 다시 수평방향으로 加算投影한다. 그림 7은 이것을 보여주고 있고, 일련의 五線 데이터와 함께 잡음성분이 존재하고 있다.

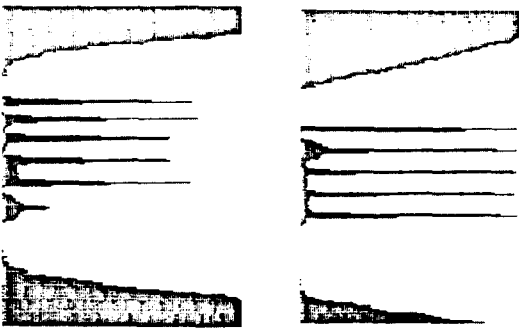


그림 7. 雜音이 있는 樂譜의 水平方向 加算技影
Fig. 7. Horizontal projections of score containing noises.

五線認識을 위해서는 우선 線을 抽出해야 하고, 抽出된 線이 5 개 이상이 되었을 때는 그중 등간격으로 되어 있는 일련의 五線을 選定하여 그것을 樂譜의 五線으로 認識한다. 이때 등간격의 허용오차는, 樂譜와 카메라와의 거리에 따라 달라지나, 한 畫面에 音標가 5

~ 8 개 있을 정도의 크기(그림 10 참조)로 했을 때, 3 畫素 이하로 한다.

이와 같이 하여 y 좌표를 추출해낸다. 또한 오선 구간 외의 모든 구간을 지워버리고 다시 수직방향 가산투영하여 오선의 x 좌표도 구할 수 있다.

3. 五線內的 마디 좌표 抽出

認識된 五線內에는 마디가 있으며, 마디는 五線의 맨 윗줄에서 맨 아랫줄로 수직으로 이은 線이므로, 認識된 五線을 윗줄부터 수직으로 加算投影하여 계속해서 연결되어지는 좌표를 그 五線에서의 마디로 認識하게 된다. 그러나 人力畫像을 CCTV 카메라로 직접 잡는 경우에는 약간의 수정이 필요하며 그 절차는 다음과 같다.

먼저, 抽出된 五線의 x, y 좌표로 人力畫像에 窓을 쬐운다. 이때 五線의 맨 윗줄 및 맨 아랫줄의 y 좌표를 각각 y_1 및 y_2 라하고, 五線의 맨 왼쪽 및 오른쪽의 x 좌표를 각각 x_1 및 x_2 라 할때, 上記 窓의 크기(그림 10(a)참조)는 x 방향에서 x_1 과 x_2 사이에 있고, y 방향에서 五線이 포함될 정도의 幅, 즉 $(3y_1 - y_2) / 2 < y < (3y_2 - y_1) / 2$ 이 되도록 한다.

그다음, 窓밖의 모든 데이터를 소거한다(그림 10(b)참조). 이때 窓 경계점에서 독립로립점의 잡음이 다시 발생할 수 있으므로 4 근방 마스크를 한번 더 쬐운다.

이렇게 窓을 설정해주므로 해서 상당량의 잡음을 없앨 수 있다. 그러나 窓內에는 아직도 잡음이 있을 수 있고, 그 양의 정도에 따라 인식에 영향을 준다. 따라서 마디좌표의 抽出條件을 다음과 같이 한다.

- ① 오선 안을 수직방향 가산투영하여 그 값이 오선간의 간격과 허용된 범위내에서 일치하는 좌표.
- ② ①에서 추출된 좌표 중에서 맨 윗줄에서 한칸 위의 공간 또는 맨 아랫줄에서 한칸 아래인 공간에서 수직방향 가산투영한 값이 정해진 값 이하인 좌표.
- ③ ①, ②의 조건을 만족하고 고립되어 있을 것.

아무리 입력화상을 깨끗이 잡아도 잡음 및 데이터 결핍현상이 생기므로, 잡음을 고려하여 투영치에 여유를 준다.

III. 認識과 演奏

1. 마디내의 各 音標의 認識과 出力 코우드의 發生
마디 안에 있는 各 音標의 認識을 위해서는 먼저 각 음표에 대해 작은 窓을 쬐운다. 이 창내에 있는 音標를 가로 및 세로 방향으로 加算投影한 값들은 그 音標의 구도에 따라 달라진다. 따라서 이 양방향의 投影值는 그 音標의 特徵을 나타내게 되고, 이들 特徵點으로 音

標의 認識을 행한다.

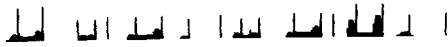
認識對象은 音樂의인 特性보다는 그들의 幾何學的인 形態에 의존하며, 다음 4 가지로 분류하여 행한다. 즉, ① 單純音標, ② 잇단음표, ③ 單純和音音標, ④ 잇단和音音標등이다.

認識過程은 다음과 같다.

① 抽出된 五線좌표와 마디좌표를 가지고서 한 마디 내를 수직방향 加算投影하여, 그 값으로 音標의 갯수를 추정한다. 이것은 音標를 머리, 대, 꼬리의 세 부분으



(a) 樂譜의 一部 (a part of score)



(b) 五線을 除外한 수직방향 加算技影 (vertical projection excluding staves)

그림 8. 수직방향 加算技影에 의해 중첩된 音標의 例
Fig. 8. Example of piled notes by vertical projection.

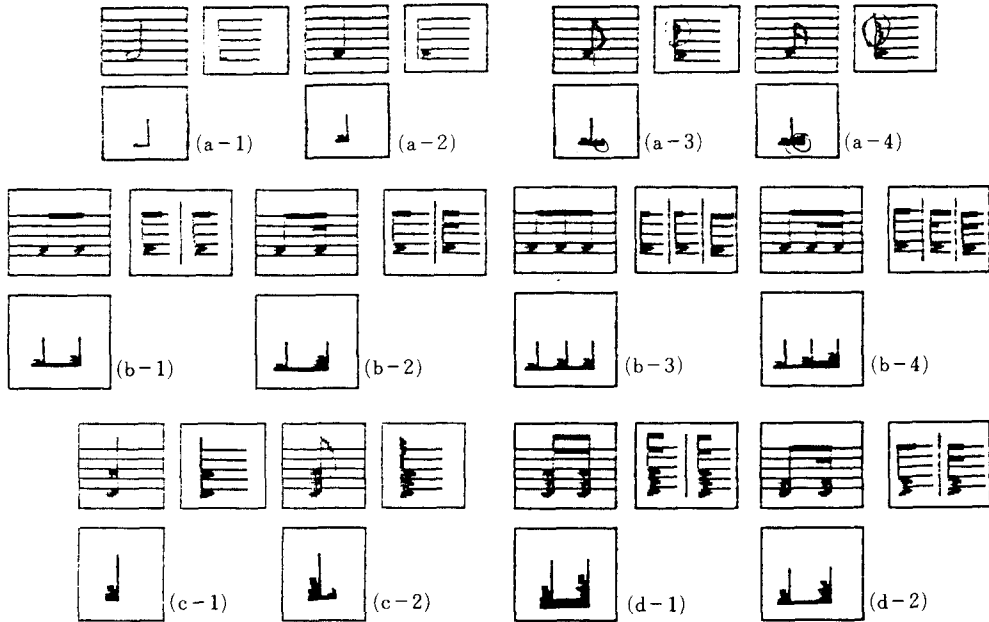
로 나누면, 온음표(O)를 제외하곤 모두 대를 갖고 있기 때문이다.

② 單純音標 및 單純和音音標, 잇단음표 및 잇단和音音標를 구분한다. 이것은 단순음표 또는 단순화음음표에 대한 認識데이터를 얻기 위해서 작은 창을 씌울 때, 한 마디에 여러개의 음표가 밀집하여 있을 경우 세로 방향 가산투영에 의해 그림8에서와 같이 2개의 음표정보가 중첩되어 마치 연결된 것처럼 보이기 때문이다. 따라서 단순음표와 잇단음표가 혼돈되어 인식될 가능성이 있게된다. 그러나 단순음표가 중첩된 경우 투영치를 보면, 두 좌표가 매우 밀집해 있거나, 또한 그렇지 않은 경우에는 두 음표 사이의 투영치가 잇단음표와 다르기 때문에 쉽게 식별할 수 있다.

③ 각각의 음표에 대해 작은 창을 씌워서 가로 및 세로 방향 가산투영치를 구하고, 오선만에 의한 가산투영치와 비교하여 각 방향에 관한 特徵벡터¹⁾를 구한다. (그림9 참조)

④ 特徵벡터는 이미 들어있는 基準패턴과 비교하여 출력코우드를 발생시킨다.

그림10은 CCTV 카메라로 잡은 화상에 대한 音標認

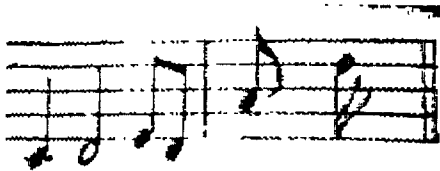


(a) 단순음표 (simple notes) (b) 잇단음표 (linked notes)
(c) 단순화음음표 (simple chord notes) (d) 잇단화음음표 (linked chord notes)

그림 9. 여러 음표의 수평, 수직방향 加算技影
Fig. 9. Horizontal and vertical projection of various notes.



(a) 入力畫像(vnput image)



(b) 窓内の 畫像(image within window)

그림10. 音標認識의 보기

Fig. 10. Example of notes recognitions.

識 過程을 보여주고 있다. 출력코우드는 음정코우드와 박자코우드로 나누어 각각에 8비트를 할당한다.

2. 출력코우드에 의한 樂器의 演奏

音標의 認識에 의해 발생된 출력코우드는 인터페이스(interface)를 통해서 컴퓨터상에서 피아노와 같은 실제의 악기로 전달되어 사람 없이 자동으로 연주가 가능하다.

그러나 본 연구에서는 컴퓨터 내장의 스피커를 이용하여 컴퓨터 자체를 악기로 만들고 있다. 스피커의 토글(toggle) 간격과 시간을 적당히 조절하면, 각 음정과 음의 길이를 만들어 낼 수 있으며, 각각의 음정과 음의 길이에 따른 서브루우턴을 만들면 우리가 원하는 음이 나오게 되어 있다.

그런데 상기 방법으로는 和音의 연주가 불가능하므로, 和音을 내는 간단한 하드웨어 시스템을 꾸미고 있다. PSG(programable sound generator)라고 하는 일종의 音樂 및 音의 合成器인 VLSI 素子를 이용하여, 그 内部 레지스터에 적당한 값을 넣어줌으로서 3중화음 8 옥타브까지 연주할 수 있게 되어 있다. 4 번째 옥타브의 도는 261.4[Hz]이며, 5 번째 옥타브의 도는 523.9[Hz]이다. 3중화음 이상의 화음을 내려면, PSG를 더 많이 부착하여야 하며, PSG 하나당 3중화음씩 증가한다. PIA(peripheral interface adapter)

로서는 6820이 6521과 호환성이 있으므로 구입이 용이한 motorola 제품인 6821을 사용하였다. 그림11에 和音演奏 시스템의 블럭도를 보인다.^{[6][7]}

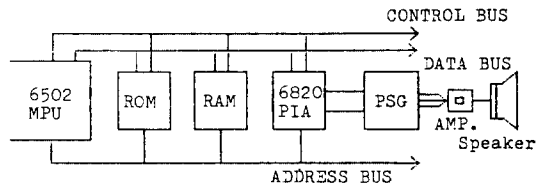


그림11. 和音 演奏시스템의 블럭도

Fig. 11. Block diagram of hardware system playing chord.

IV. 結論 및 檢討

印刷 樂譜 畫像을 직접 CCTV 카메라로써 컴퓨터에 人力시켜 이 畫像을 認識, 演奏하는 시스템에 관해 논했다. 이 認識의 過程에서 認識率을 決定지워 주는 要素들은, ① 認識 알고리즘, ② 周圍의 照明狀態, ③ 樂譜의 鮮明度 및 기울림 狀態, ④ 樂譜와 CCTV 카메라와의 距離, ⑤ CCTV 카메라의 解像度, ⑥ 컴퓨터 그래픽 解像度, ⑦ 비디오 信號 디지털이저의 性能등이라 할 수 있다. 여기서, ⑤⑥⑦은 시스템에 의해 決定되어 버리는 것이며, 認識에 주로 關여하는 것은 ①②③④라고 볼 수 있다. 이중 實際의인 側面에서 ②④는 認識率을 크게 변화하게 했다. 특히 照明狀態는 畫質을 좌우했고, 認識 알고리즘(加算投影法)의 性質로 비스듬히 촬영된 樂譜 혹은 데이터 결핍이 심한 경우는 認識에 실패했다. 그러나 雜音이 전혀 없는 樂譜의 경우, 五線, 마디, 音標를 認識對象으로 한 本 研究에서, 認識에 실패하는 일은 거의 없었다. 보통의 형광등 照明狀態라 하더라도 필터를 사용할 경우, 약 90% 이상의 認識率을 얻을 수 있었다. 本 實驗에 사용된 CCTV 카메라 및 디지털이저의 性能이 과히 좋지 않은 상태였으므로 이 부분을 개선한다면 認識率은 보다 向上될 수 있으리라 생각된다. 온음표는 대가 없기 때문에 부호 인식의 일부분으로 간주하여 다음으로 미루었다.

처음의 연주에서 실시간 연주가 되어지고 있지 않다. 이는 PSG에 타이머를 연결하면 실시간 연주도 가능할 것이다. 아울러 窓内に 雜音이 많은 畫像 혹은 데이터 결핍이 많은 화상 등에 관한 연구도 있어야 할 것이다.

거의 모든 프로그램은 6502 어셈블리어로^[8] 작성하여 실행 속도를 높였다. 그림1.2는 전체적인 시스템 블럭도이다.

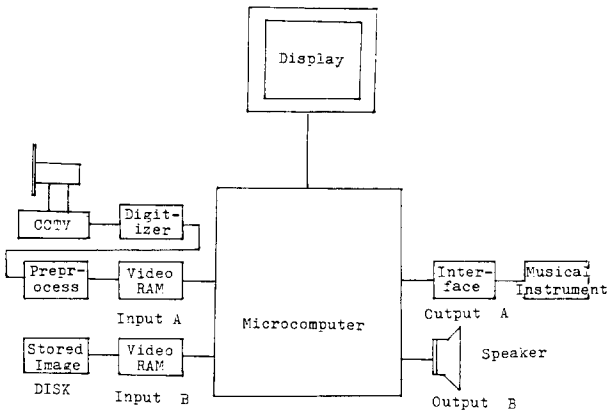


그림12. 전체적인 시스템 블록도
Fig. 12. Total system block diagram.

參 考 文 獻

[1] R.C. Gonzalez, P. Wintz: *Digital Image Processing*. Addison Wesley, 1979.
[2] D.H. Ballard and C.M. Brown: *Computer*

Vision, Prentice-Hall, 1982.

- [3] R.O. Duda and P.E. Hart: *Pattern Classification and Scene Analysis*. John-Wiley & Sons, 1973.
[4] 安居院猛, 中嶋正之: *コンピュータ画像処理*, pp. 65-75, 産報出版社, 1980.
[5] 坂井利之外: “計算機による顔写真の解析” *電子通信學會論文誌*, 56-D, 4, pp. 226-233 1973.
[6] Rodney Zaks: *6502 Application Book*. Sybex, 1983.
[7] J.D. Greenfield, W.C. Wray: *Using Microprocessor and Microcomputers: The 6800 Family*, Wiley, 1983.
[8] Rodney Zaks: *Programming the 6502*, Sybex, 1983.
[9] 이명우, 최종수: “컴퓨터를 이용한 인쇄악보의 인식과 연주(2)”, *전자공학회추계종합학술대회논문집*, vol. 6, no. 2, pp. 187-190, 1983.
[10] 이명우, 최종수: “컴퓨터를 이용한 인쇄악보의 인식과 연주(3)”, *전자공학회 추계종합학술대회 논문집*, vol. 7, no. 1, pp. 357-360, 1984.