**Advanced Computer Vision**

Term Project

R03922126 蘇兆為

**Title**

Automatic Music Score Recognition

**Introduction**

The objective of this project is to automatically recognize music score, no matter from the images from scanner, or from a simple mobile phone camera. After music score information is extracted, we can use these data to create the MIDI file and play this music file.

**Related work**

1. 鋼琴樂譜辨識以及撥放系統 Piano Score Recognition and Playback System(2011)

<http://140.134.131.145/upload/paper_uni/992pdf/%E9%8B%BC%E7%90%B4%E6%A8%82%E8%AD%9C%E8%BE%A8%E8%AD%98%E5%8F%8A%E6%92%AD%E6%94%BE%E7%B3%BB%E7%B5%B1.pdf>

1. 印刷樂譜辨識系統Automatic Recognition of Printed Music Score(2004)

<http://image.cse.nsysu.edu.tw/2004student/Wei/Automatic%20Recognition%20of%20Printed%20Music%20Score.pdf>

The former uses simpler method for recognition. And the latter one is more complicated, and the result is better. I will mainly refer to the latter one, and try to improve it.

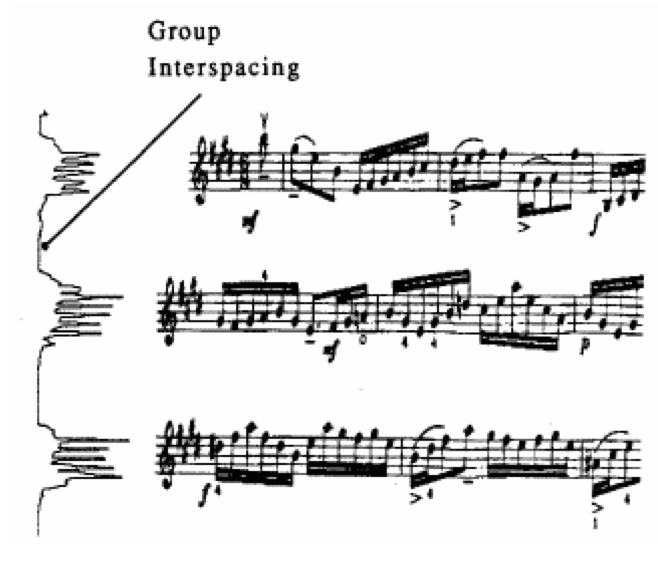
**Method**

* RGB to binary

First, the music score is only black and white image. So if our input images contain RGB channels, convert it to gray level image, then use threshold to do binarization.

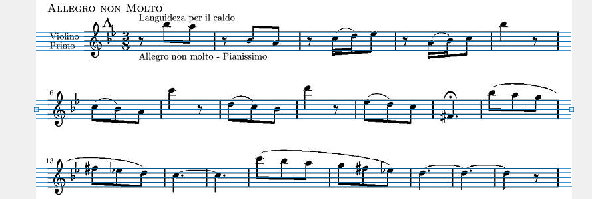
* Score line detection

Score lines play an important role in locating the music notes, so we need to find it out. Hough transform is a powerful way to do straight line detection, and it might be useful in this work. But if our input images are good enough, the score line should be horizontal, and we don’t need such powerful method. We can simply use orthographic projection to detect the score lines. First, project the image to y-axis. We can do this simply by counting the number of pixels every row, then draw the histogram of result. It should be like this:



Result of orthographic projection

We can easily find that the peaks of histogram occur in the score lines position, so theoretically we can just find the five peaks per group, then we done. The problem is, assume we can find the group position without confusing with other group, what is the threshold that is used to distinguish whether is peak? If the threshold is not chosen properly, we might detect more than or less than five lines in the group. The second problem is that because the score lines are not always one pixel width, the top 5 peaks are not always the different lines. To fix these problems, we need to exclude the peaks that are too close to the others or too weak, then pick the top 5 peaks. If we can’t find out enough lines, we need to use the lines that have been found to calculate the space between two lines, and recover the missing lines.



Result of line detection

After we detect the score lines, we can begin to detect the music notes. But before that, we need to remove the score lines from the image first. We can assume the line width is about 2 or 3 pixels, and when we remove the lines, we should not destroy the music notes so that the note detection can be easier.



Result of line removal

**Expected Result**

We will use the images from scanner for experiment first, because it would be clearer, and less distortion. Then begin to use images from mobile phone camera. We can even try to take these images obliquely on purpose. Finally, we hope this system can be applied to handwritten music score.