Symbol Recognition of Printed Piano Scores with Touching Symbols

Fubito Toyama, Kenji Shoji, and Juichi Miyamichi
Faculty of Engineering, Utsunomiya University
7-1-2, Yoto, Utsunomiya-shi, 321–8585 JAPAN
fubito@is.utsunomiya-u.ac.jp

Abstract

To build a music database efficiently, an automatic score recognition system is a critical component. Many previous methods are applicable only to some simple music scores. In case of complex music scores it becomes difficult to detect symbols correctly because of noise and connection between symbols included in the scores. In this paper, we propose a score recognition method which is applicable to the complex music scores. Symbol candidates are detected by template matching. From these candidates correct symbols are selected by considering their relative positions and mutual connections. Under the presence of noise and connected symbols, the proposed method outperformed "Score Maker" which is an optical music score recognition software.

1. Introduction

Optical music recognition (OMR) is automatic recognition of a scanned music score. OMR is a critical component to build a music database efficiently. Converting printed scores into computer readable form can be applied to edit music scores easily. Furthermore, there is a great demand for OMR from users who cannot read music scores.

Work on OMR began in the late 1960's and there are many works on OMR [1]–[5]. A lot of commercial OMR software exists today, but it does not have enough performance compared with character recognition. Although it seems to be a rather simple problem because printed music score is constructed by simple symbols, it is in fact much more difficult than it seems to be. The main sources of difficulty are segmentation problems such as two touching symbols when syntactically they should not (e.g. connection between a sharp and a note head). Therefore, it is difficult to recognize complex music scores in which symbols are drawn with high density, although existing recognition systems achieve good recognition rates for simple music scores. In traditional recognition systems, first, staff lines are detected. After that their staff lines are removed. Next, the symbols are segmented using connected component analysis, and they are detected. Thus, symbols cannot be detected if normally separated symbols touch each other. For this problem, music-syntactic rules and knowledge are generally used. Couasnon and Retif [4] proposed a grammar to formalize the musical knowledge.

In this paper, we propose an OMR system for the complex music scores which contain many touching symbols and noise. In many cases, scanned music scores contain noise caused by a binary image threshold and poor paper quality. But these scores are out of scope of previous works. There are many touching symbols and shape deformation in these scores. Our method handles these scores. For detecting these symbols, symbol candidates are detected by template matching with a low threshold. From detected these candidates, correct symbols are selected by checking their relative position and mutual connections. In our experiments, recognition symbols (primitives) are note head, stem, beam, hook, natural, sharp and flat. The proposed method outperformed "Score Maker" which is an optical music score recognition software.

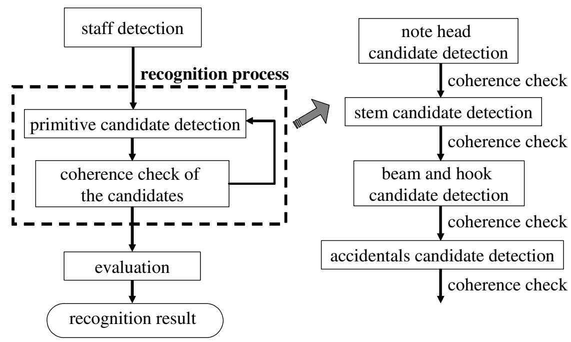

Figure 1. The processing flow.

2 General overview

From scanned binary images, our system recognizes the primitives such as note head (black and white head), stem, beam, hook, natural, sharp and flat. Figure 1 shows the processing flow. First, staff lines are detected. The traditional methods [1][5] using vertical projection profile are used for staff line detection in our method. The recognition process is divided into two steps. The first one is the primitive candidate detection. The second one is the coherence check of these candidates. This process starts from note head detection and is performed for each primitive, because there are many note heads and many other primitives are written based on their position. Stem detection is performed using the positions of note head candidates. In the coherence check of the stem candidate detection, the check of note head is also performed again. The check of the primitives which had been detected are performed again in the each

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

process of the primitive detection.

In the primitive candidate detection, candidates are detected by pixel tracking and template matching with a low threshold. This process allows to detect incorrect primitives (false positive), but it is necessary to detect all correct primitives. In the coherence check, the touching positions between the primitives are estimated by checking the contour positions of the primitives. Finally, the contour positions and relative positions between the primitives are evaluated by referring to music writing rules, and the symbols are recognized. The evaluated symbols are note head and stem in our present system.

3 Staff detection

Staff lines contain important information such as scale of primitives and image skew. Therefore, first, staff lines are detected and other primitives are detected based on the staff lines. The traditional methods [1][5] using vertical projection profile are used for staff line detection in our method. First, run length histograms of the black and white pixels are constructed from equally spaced vertical scan lines. Staff candidate points are found from the histograms. Staff lines are detected by finding black pixels on lines which are linked to the candidate points.

4 Recognition process

As mentioned in Section 2, the recognition process is divided into two steps, and it starts from note head detection. In the first step, candidates of a primitive are detected. In the next one, contour positions and relative positions between the candidates are checked. This process is performed for each primitive. The coherence check of the primitives which had been detected is performed again in the each process of the primitive detection. The checked primitives increase as the process advances. The coherence check is described in detail in Section 4.1.2.

4.1 Note head detection

In music scores, there are many note heads, and their positions are estimated from the position of staff lines. Furthermore, many other primitives are written based on note head position. Thus, first, note heads are detected. Candidates of note heads are detected using template matching with a low threshold. In the coherence check, the presence of the touching primitives is decided by checking the contour positions of the candidates. Black and white note heads (whole note heads excepted) are detected in our experiment.

4.1.1 Detection of note head candidates

Candidates of the note heads are detected using template matching with a low threshold. We use three types of templates, black note head, white note head on staff lines and on staff spacing, respectively. Note heads are written on staff lines, staff spacing, ledger lines, spacing between ledger lines, and spacing between a ledger line and a staff line.

(a) tracking to top and bottom (b) tracking to left and right

Figure 2. Contour detection of note heads.

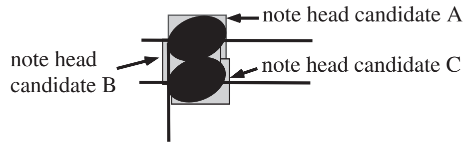

Figure 3. Example of the coherence check.

Therefore, template matching is performed only in these regions. Ledger lines are written above and below the staff, and their intervals are roughly equal to staff spacing. Ledger lines are detected by searching horizontal run length in their regions. If the ledger lines are detected, the regions above (or below) the ones are searched again. Note head candidates are detected at the positions in which matching rate is higher than a threshold. If their positions are close to each other, the position having the highest matching rate is selected.

4.1.2 Coherence check of note head candidates

The touching primitives are estimated by checking the contour positions of the candidates. Whether there are the touching primitives or not is decided by referring to music writing rules.

In case of note heads, the contour positions are detected by tracking black pixels to top, bottom, left and right as shown in figure 2. The contour points are classified into three classes as follows.

Class 1: the contour position is correct. The contour position is within the range of the primitive one.

Class 2: the contour position is incorrect. The contour position is out of the range of the primitive one.

Class 3: the contour position is incorrect, but is touching other primitive. In case of class 2, there is other primitive near the contour position, and their relative positions are correct on music writing rules.

In cases of tracking to left and right, the contours on the staff lines are excepted from the coherence check. For example, in a note shown in figure 3, three candidates (A, B and C) of note heads are detected by template matching. The candidate B is detected as a incorrect note head because of a template with a low threshold. In the coherence check, the candidate C is decided as touching the candidate A. Similarly, the candidate A is decided as touching the candidate C. Thus, in the candidate A and C, the contours classified as class 2 change to class 3. On the other hand, in the candidate B, there is no touching candidate on the estimated contours. Thus, the contours classified as class 2 do not change.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

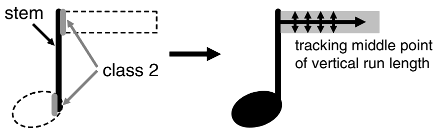

Figure 4. Beam detection.

4.2 Stem, beam and hook detection

4.2.1 Detection of stem candidates

A stem is generally attached to the top right or bottom left of a note head. Thus, the stem candidates are searched by tracking black pixels to the top (bottom) from the top right (bottom left) of note head candidates. If there is connection of note heads because of chord description, the same stem candidate is detected repeatedly. In this case, the longest stem candidate is selected. The region in which run length is larger than $2S_i$ (S_i is staff spacing) becomes the stem candidate.

4.2.2 Coherence check of stem candidates

In the coherence check of the stem candidates, the right and left side contour positions are checked and classified into three classes as described in Section 4.1.2. The contours on the staff lines are excepted from the coherence check, as well as the case of the note head. If there are continuous contour points which are classified as class 2, there is a possibility to touch another primitive. Therefore, beam and hook are searched from this region because they always connect to the stem.

4.2.3 Detection of beam and hook

Beam and hook are searched from the region which is classified as class2 in the coherence check of stem candidates. Beam is detected by tracking the middle point of vertical run as shown in figure 4. The points which have the run in the range of beam height are tracked. Beam is detected by the relative position from these points. Similarly, the points which have the run in the range of hook height are tracked, and the hook is detected by the relative position from these points.

4.3 Detection of accidentals

Flat, sharp and natural can be detected in our present system. These primitives are detected using template matching. The common templates cannot be used because the shapes of these primitive are different for each score. Therefore, the templates are made from primitives in the score. The accidentals without noise are selected as these primitives. The accuracy of these templates is high when many accidentals are selected. Each template is made by calculating the average of these primitives. The other accidentals with noise are detected by template matching with a low threshold.

The method of making templates is detailed as follows. The accidentals are searched in the left side region of note head candidates, because accidentals except key-signature are generally written on the left side of a note head. In this region, bar lines included in the accidentals are detected tracking black pixels to top and bottom. Furthermore, features shown in figure 5 are detected using the bar lines. Natural is detected from widths of bar A and B, endpoints C, D, E and F shown in figure 5(a). Similarly, sharp is detected from widths of bar A and B, endpoints C, D, E and F shown in figure 5(b). Flat is detected from run length histogram of A shown in figure 5(c). This histogram is calculated in the range B. The templates are made by calculating the average of images of each detected accidental. Using these templates, the other accidentals are detected by template matching with a low threshold. Template matching is performed in the left side region of note head candidates.

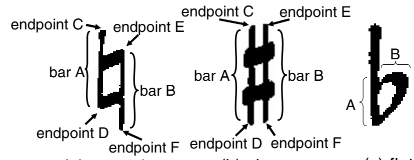

Figure 5. Detection of accidentals.

5 Primitive evaluation

Each primitive is evaluated using the classified classes of its contours. The coherence check of note heads and stem is implemented in our present system. Implementation of the coherence check of other primitives is future work. In the note head evaluation, the rate of classified class 1 and 3 in the contour points of a note head candidate is calculated. Each note head candidate is evaluated by this rate and the evaluation value of the stem candidate attached to its candidate. Finally, note heads are detected from this evaluation value. Similarly, the rate of classified class 1 and 3 in the contour points of a stem candidate is calculated. Stems are detected from this rate.

6 Experimental results

Two types of printed piano scores are used in our experiments. One is the noiseless images (score set A) scanned from printed piano scores. The others are the images (score set B) which contain a lot of noise caused by poor paper quality. Furthermore, score set A and score set B are classified into two types of scores in which symbols are drawn with high density (A1,B1) and low density (A2,B2). Thus, four types of scores are used in our experiments. In order to reduce noise, the binary threshold value is set interactively when a scanned image is transformed into a binary image. The parts of the scores A1 and B2 are shown in figure 6 and 7, respectively. Figure 8 shows magnification of note part of figure 7. The score B2 contain a lot of noise as shown in figure 8. Recognition primitives are note head, stem, beam, hook, natural, sharp and flat in our experiments.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Table 1. recognition results (proposed method)

score	black head	white head	stem	beam	hook	accidentals	# of false positive	recognition rate
A1	343/344	51/51	194/194	2/2	42/42	30/30	1	99.70 %
A2	516/517	3/3	361/361	67/67	0/0	126/126	1	99.81 %
B1	584/595	17/21	506/519	237/272	23/41	14/19	23	92.57 %
B2	469/478	5/5	423/441	236/276	3/7	4/13	7	92.87 %

Table 2. recognition results (Score Maker)

score	black head	white head	stem	beam	hook	accidentals	# of false positive	recognition rate
A1	344/344	51/51	194/194	2/2	42/42	30/30	0	100 %
A2	515/517	3/3	359/361	67/67	0/0	123/126	0	99.35 %
B1	432/595	7/21	349/519	157/272	41/41	12/19	5	67.69 %
B2	448/478	5/5	421/441	238/276	3/7	10/13	12	91.23 %

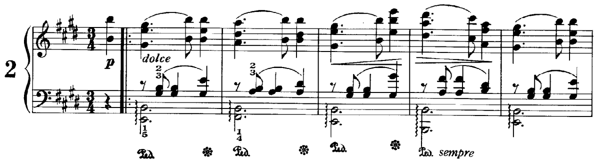

Figure 6. Part of score A1.

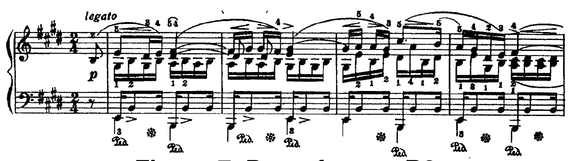

Figure 7. Part of score B2.

Recognition results are shown in table 1. In table 1, the denominator represents the number of primitives in the score, and the numerator represents the number of correctly recognized primitives. The recognition rate is defined as

$$Recognition\ rate = \frac{N_{rec} - N_{fp}}{N_{all}} \qquad (1)$$

where N_{rec} is the number of correctly recognized primitives, N_{fp} is the number of false positives, and N_{all} is the number of primitives in the score. We achieve high recognition rates for score A1 and A2. Although the recognition rates of score B1 and B2 are lower than those of score A1 and A2, we can achieve good recognition rates. There are many false negatives for the beam of score B1 and B2. The main cause is that the coherence check of the beam is not implemented in our present system. These recognition rates will be improved by implementing the coherence check of other primitives. The false negative rate of the hook is low for score B1, because the shape of hooks in score B1 is different from standard ones. Therefore the method of the hook detection needs to be modified.

We compared our method with Score Maker which is a optical music score recognition software. Table 2 shows the recognition results of Score Maker. The recognition rates of score A1 and A2 is high, but these of score B1 and B2 are low in table 2. As seen in table 1 and 2, in the score set A

Figure 8. Magnification of figure 7.

(noiseless images), recognition performance of our method was almost the same as that of Score Maker. On the other hand, in the score set B which contains a lot of noise, the proposed method outperformed Score Maker. Especially, the paper quality of score B1 is very poor. The proposed method is useful for such scores.

7 Conclusion

In this paper, we have proposed a score recognition method which is applicable to the complex music scores. We have applied template matching and coherence check to detect touching primitives. Candidates of the primitives are detected by template matching. In the coherence check, correct primitives are selected by checking contour position of these candidates. In our experiments, the proposed method is useful for the complex scores which contain noise, and outperformed Score Maker.

References

[1] H. Miyao and Y. Nakano. Note symbol extraction for printed piano scores using neural networks. *IEICE Trans. Inf. & Syst.*, vol. E79-D, no. 5, pp. 548–554, May 1996.

[2] N. P. Carter and R. A. Bacon. Automatic recognition of printed music. *Structured Documnet Image Analysis*, pp. 456–465, 1992.

[3] F. Rossant. A global method for music symbol recogition in typeset music sheets. *Pattern Recognition Letters*, vol. 23, pp. 1129–1141, 2002.

[4] B. Couasnon and B. Retif. Using a grammar for a reliable full score recognition system *Proc. ICMC*, pp. 187–194, 1995.

[5] T. Matsushima, T. Harada, I. Sonomoto, K. Kanamori, A. Uesugi, Y. Nimura, S.Hashimoto and S. Ohteru. Automated recognition system for musical score – the vision system of WABOT-2 – *Bulletin of Science and Engineering Research Laboratory, Waseda University*, no. 112, pp. 25–52, 1985.