Staff Line Detection by Skewed Projection

Diego Nehab

May 11, 2003

Abstract

Most optical music recognition systems start image analysis by the detection of staff lines. This work explores simple techniques from document image analysis, such as line segment extraction, to guide the staff line identification process. Specifically, the overall document skew is computed from the detected line segments. Staff lines are then projected in the direction of the detected skew and identified as maxima in the histogram of projections. The method is simple to code and robust even for unrealistically large rotations of the input data.

1 Introduction

Optical Music Recognition (OMR) is the process of automatically recovering the information present on music scores based on scanned data. The goals and applications are similar to those of Optical Character Recognition (OCR). A working OMR system would be helpful in many ways. It could be used, in conjunction with a synthesis system, to play music directly from printed scores. Old music documents could be automatically converted to any of several digital formats for later reprinting in higher quality typesetting or storage in digital libraries. Recognized music could be converted to braille notation or transposed and printed again. For reviews of previous work in OMR and further applications, see [3, 2, 1].

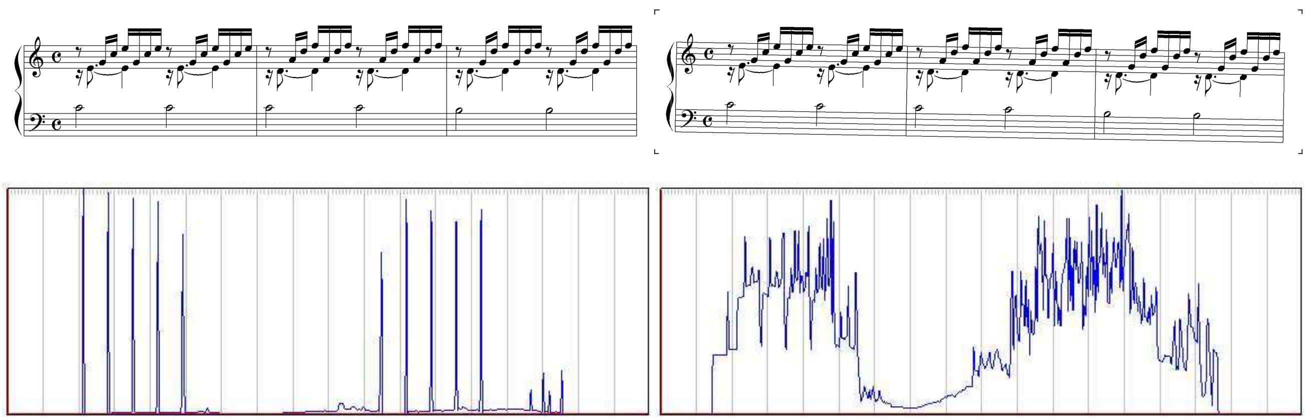
Because staff lines are the dominant feature in music scores and must be detected at some point, most OMR systems start the recognition process with staff line identification. Not only do staff lines provide a vertical coordinate system for musical primitives, but their spacing and thickness also provide the basic scale for relative size comparisons. Although the presence of staff lines is helpful for these reasons, it is also harmful. Staff lines usually connect several independent symbols, may fill otherwise empty symbol regions and completely cover other symbol features. For that reason, many OCR systems eliminate staff lines before the recognition phase.

In real data, scanned from music documents, staff lines are not always horizontal, straight, contiguous, equidistant, do not have constant thickness and are not even parallel. Therefore, a robust staff line identification method should make as few as possible of the above assumptions. Several methods have been proposed, each with its strengths and weaknesses. Our method assumes staff lines are parallel and straight. A brief review of previous work is given in the following section.

2 Previous methods

A common approach to staff line identification is known as *horizontal projections* [3]. The idea is to compute a histogram of the number of black pixels along each row of the digitized music score. Rows corresponding to staff lines should be easily identified as peaks in the histogram. Although very simple, this method is also very sensitive to skew. Figure 1 shows the resulting horizontal histogram for part of a Bach's piece. Figure 1(a) is perfectly horizontal whereas 1(b) was rotated by only one degree. It is unclear whether it is even possible to identify the staff lines by the histogram of the rotated image. The fact that scanned documents are often skewed by greater amounts renders the method almost useless. To overcome this shortcoming, several other algorithms have been developed. Most of them achieve robustness by building staff lines from localized features.

Kato and Inokuchi [9] method first detects staff line thickness and separation as the maxima of histograms of the run-lengths of black and white pixels in image columns, respectively. This information is used to estimate the size of a window that will be large enough to contain a complete staff. Windows are placed in the leftmost



(a) (b)

Figure 1: Histogram of horizontal projections.

and rightmost corners of each staff. Horizontal projections on each window are used to find staff line vertical positions and vertical projections are used to find staff line horizontal boundaries.

Randriamahefa et al. [11] search for regions of minima in the vertical projection of the image. These regions are assumed to contain only staff lines. For each region, several local horizontal projections are computed. Wide peaks are rejected as possible beam parts and thin peaks are declared part of staff lines. The detected staff line segments are later combined into complete staff lines by the fit line square method.

Carter and Bacon [4] algorithm converts the music image into its Line Adjacency Graph (LAG) representation. The image is scanned from top to bottom and vertical run-lengths are stored for each column. Then, adjacent vertically overlapping run-lengths are horizontally linked together. To detect staff lines, candidates are selected from *filaments* (high aspect ratio segments of the LAG) with constrained inclination and thickness.

Leplumey et al. [10] also compute vertical run-lengths and group them horizontally. Groups are first formed where the thickness is similar to the staff line thickness estimated by the most common vertical run-length. These are called *dashes*. The remaining horizontal groups are called *stains*. Each dash becomes a vertex in a graph on which vertices are linked if their corresponding dashes are close together and have approximately the same orientation. The graph is then searched for groups of five long connected components, which are classified as a staff.

3 Going vectorial

Bainbridge and Bell [1] break OMR into four smaller steps: *staff line identification*, *musical object location*, *musical feature classification*, and *musical semantics*. The first two steps isolate the musical primitives that are grouped, classified and interpreted by the last two steps. Therefore, after the first two steps, we should be left with a vectorial representation of the music score, sufficient perhaps to reprint the score at higher quality, but with no associated musical meaning. It might be easier to initially convert the music score data into vectorial form and only then start analyzing it. This is the main idea behind the present work, and is responsible for the new method for staff line detection presented in the next section.

Many methods have been proposed to extract vectorial information from digital images. The research in the area is motivated mainly by the need to recognize flow charts, electronic circuits and similar diagrams. For a good overview of the available tools, see [8]. In order to keep things simple in the first attempt to perform staff line recognition, vectorization was performed in three easy steps: *thinning*, *segment extraction*, and *polygon simplification*. All steps are reviewed in the following sections.

3.1 Thinning

In the first step, the input image is reduced to its medial axes. The simplest method, known as Hilditch skeletonization [6], was used. It proceeds by progressively eroding the objects, while carefully preserving the original topology. When no further pixels can be eroded, the algorithm halts.

Figure 2 shows the products of the process. The gradient shows the round at which each pixel was eroded and provides a measure of the distance from each object pixel to the background. The object pixels that remain after the last round are shown in the rightmost image.

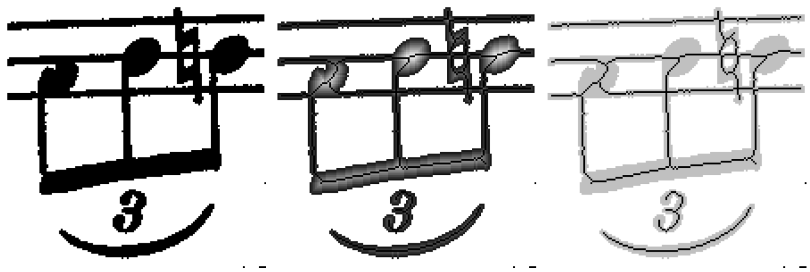


Figure 2: Hilditch skeletonization.

Naturally, there are many other skeletonization techniques [14]. It would be interesting to analyze some of them, specially those that produce better results at intersections [7] or are more stable with respect to rotations [13].

3.2 Segment Extraction

The segment extraction starts after skeletonization, with the identification of *critical points* (skeleton joints and tips). These are all pixels in the skeleton that do not have exactly two 8-neighbors. The critical points are erased from the skeleton so that only simple disconnected segments are left. These independent segments are then collected near the critical points by a series of simple flood fills. Figure 3 shows the process.

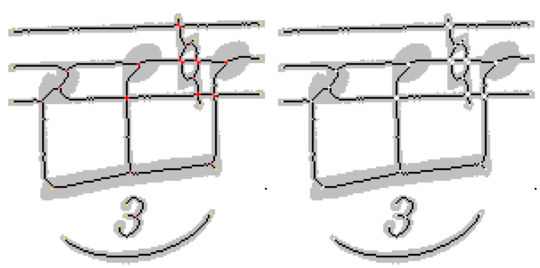


Figure 3: Segment extraction.

In the current implementation, critical points are not added as parts of the segments. This prevents the simplification process from taking them into account. Naturally, this limitation should be removed in the future.

3.3 Polygon Simplification

During segment extraction, each individual segment is collected as a sequence of points, just like a polygonal line. Applying polygon simplification ensures that the number of vertices in each segment is as small as possible, without altering too much the shape of the segment. The simplification is specially important for the later stages of staff line detection, which analyze the orientation of lines between adjacent segment vertices.

Again, there are many methods for polygonal approximation [12]. The method chosen for the current implementation is known as Douglas-Peucker polygon simplification [5]. It starts by marking the polygon end-points as part of the simplified version. Then, it searches for the vertex in the original polygon that is furthest from the line segment linking the marked endpoints. If the distance is greater than some fixed threshold, the furthest vertex it is added to the simplification and the algorithm proceeds recursively in the left and right sub-polygons defined by the end-points and the newly selected vertex. If no vertices are further away than the threshold, the algorithm halts.

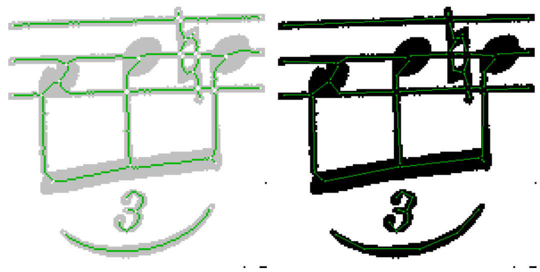


Figure 4: Polygon simplification.

Figure 4 shows the result of the simplification. In order to help the skew detection, a large threshold is used during simplification, to get rid of noise and grid biases.

4 Staff line detection

The detection of staff lines is done in three steps. First, the overall skew in the scanned image is detected from the extracted line segments. Then, a projection histogram, the *skewed projection*, is computed in the direction of the previously found skew. Finally, staff lines are found as the maxima in the skewed projection histogram.

4.1 Skew detection

Since the staff lines are the predominant feature in the scanned data, we can expect that the staff line orientation will be the most common segment orientation. For that reason, the overall image skew can be detected as the maximum in the segment orientation histogram. Figure 5 shows such a histogram, computed from the rotated score of Figure 1. Notice the detected peak at exactly one degree.

Figure 5: Orientation histogram.

In the current implementation, the histogram precision is one tenth of a degree and, before peak detection, a low pass filter is applied to it in order to remove noise and grid biasing. The table below shows the detected and measured skews for the test samples. The results are fairly precise.

	bach	**rach**	**hanon**	**czerny**
measured	-1.00	1.85	-5.82	-5.61
detected	-1.0	1.9	-5.9	-5.5

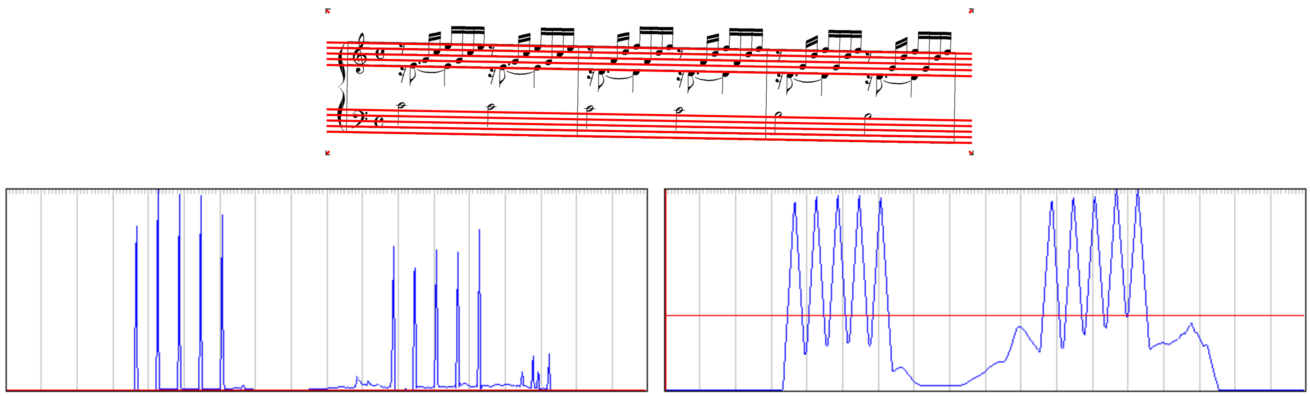
4.2 Skewed projection

Given the staff line skew, it is possible to either rotate the line segments to the horizontal orientation or to directly project them in the direction of the skew. Projecting in the skew direction is very simple and also eliminates the confusion caused by rotation. The resulting skewed projection histogram is as clean as the horizontal projection histogram computed from perfectly horizontal images.

Figure 6(a) shows the skewed projection histogram for the rotated Bach's score. It is clear that the confusion has been eliminated. It is now possible to detect the maxima corresponding to the staff lines.

4.3 Maxima detection

Again, before maxima detection, the histogram is low-pass filtered. Then, proceeding from the topmost value, a horizontal line is swept downward and intersections with the histogram polyline are computed. The sweep is halted as soon as the number of intersections start to decrease[1]. At that point, all midpoints between upward and downward intersections (in this order, of course) are reported as staff line positions. Figure 6(b) shows the smoothed histogram and the position where the sweep line halted. The peaks correspond to the staff lines shown in the score image.



(a) (b)

Figure 6: Skew projection and detected staff lines.

5 Conclusions and Future Work

This work explored the information obtained from a simple line segment extraction method to guide the projection histogram used for staff line extraction. The contributions were a method to detect the overall image skew and the application of this knowledge in the construction of a skewed projection histogram from which the staff lines could be easily found. Considering how simple the method is, the results are very satisfactory.

However, the original idea was to use connectivity information between the detected line segments to guide a graph based algorithm that would find the longest straight lines. The information provided by connectivity would help prevent the false positives seen in Figure 9 (in green). These were caused by the great number of ledger lines and beams aligned with the staff line orientation. The false positives caused by text (in blue) could be avoided by removing the text prior to staff line detection (using any of many available OCR techniques). Also, knowing exactly which line segments belong to the staff lines would be important for its removal, and a graph based method would provide that information immediately. Finally, the graph based method could be more robust to distortions in the scanned image, being able to detect staff lines that are not parallel or straight.

Finally, it would be also interesting to investigate other techniques for segment extraction, specially higher quality skeletonization and polygon simplification.

[1]Thanks to Szymon for the idea.

References

[1] D. Bainbridge and T. Bell. The challenge of optical music recognition. *Computers and the Humanities*, 35 (2), 2001.

[2] D. Bainbridge and N. P. Carter. Automatic reading of music notation. In H. Bunke and P. S. P. Wang, editors, *Handbook of Character Recognition*, pages 583–603. World Scientific, 1997.

[3] D. Blostein and H. S. Baird. A critical survey of music image analysis. In H. S. Baiard, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 405–434. Springer-Verlag, 1992.

[4] N. P. Carter and R. A. Bacon. Automatic recognition of printed music. In H. S. Baiard, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 456–465. Springer-Verlag, 1992.

[5] D. Douglas and T. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartografer*, 10(2):111–122, 1973.

[6] C. J. Hilditch. Linear skeletons from square cupboards. In *Proceedings of the 4th Anual Machine Intelligence Workshop*, volume 4, pages 403–420. Edinburgh University Press, 1969.

[7] G. Hu and Z.N. Li. An x-crossing preserving skeletonization algorithm. *International Journal of Patternt Recognition and Artificial Intelligence*, 7(5):1031–1053, 1993.

[8] R. Kasturi, R. Raman, C. Chennubhotla, and L. O'Gorman. An overview of thechniques for graphics recognition. In H. S. Baiard, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 285–324. Springer-Verlag, 1992.

[9] H. Kato and S. Inokuchi. A recognition system for printed piano music using musical knowledge and constraints. In H. S. Baiard, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 435–455. Springer-Verlag, 1992.

[10] I. Leplumey, J. Camillerapp, and G. Lorette. A robust detector for music staves. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 902–905. IEEE, 1993.

[11] R. J. Randriamahefa, J. P. Cocquerez, C. Fluhr, F. Pépin, and S. Philipp. Printed music recognition. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 898–901. IEEE, 1993.

[12] P. L. Rosin. Techniques for assessing polygonal approximations of curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):659–666, 1997.

[13] G. Sanniti di Baja. Well-shaped, stable, and reversible skeletons from the (3,4)-distance transform. *Journal of Visual Communication and Image Representation*, 5:107–115, 1994.

[14] C.Y. Suen and P.S.P. Wang, editors. *Thinning Methodologies For Pattern Recognition*, volume 7 of *International Journal of Patternt Recognition and Artificial Intelligence*. 1993.

Figure 7: Rachmaninoff.

Figure 8: Czerny.

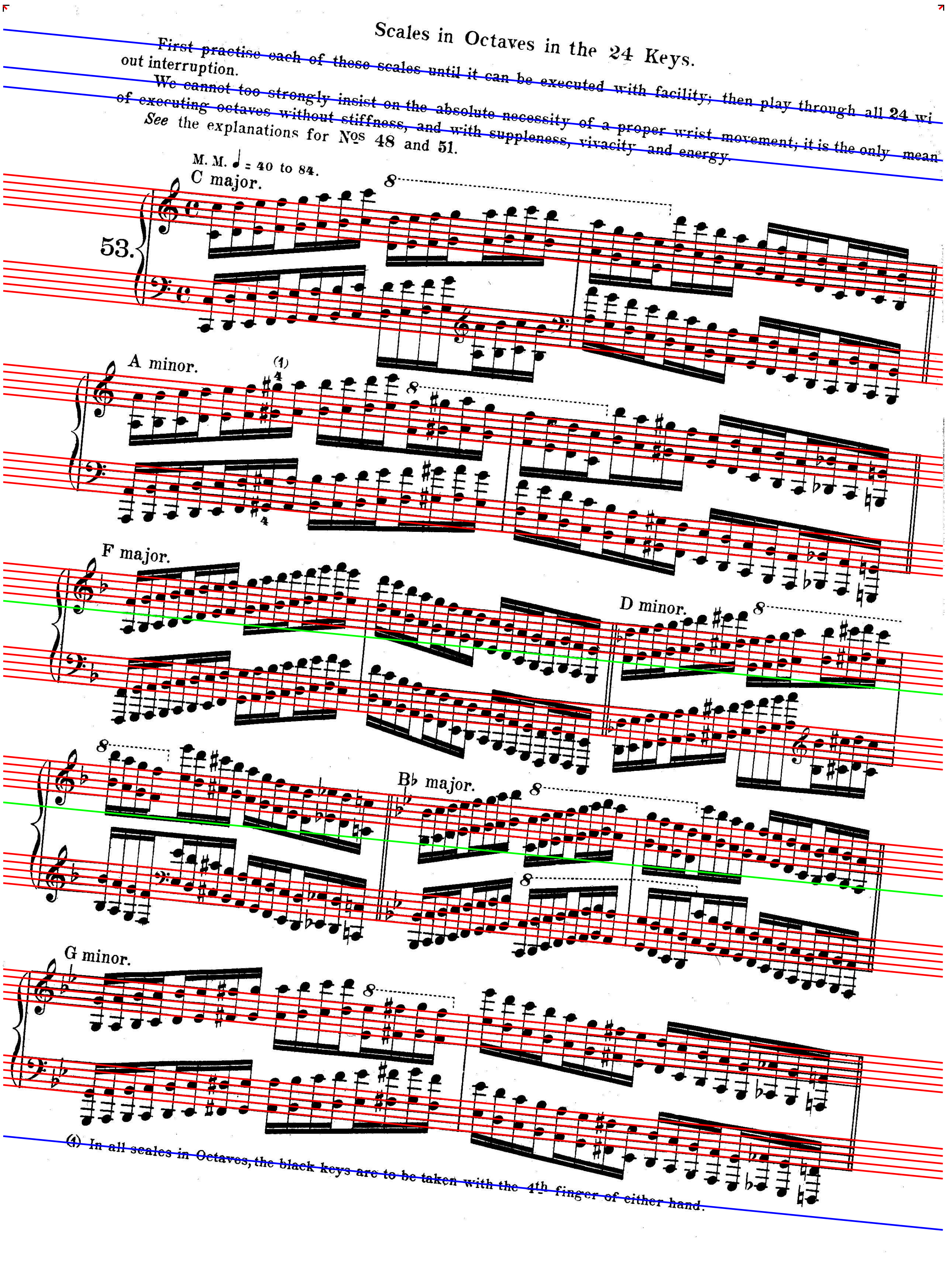


Figure 9: Hanon.