A Robust Detector for Distorted Music Staves*

Mariusz Szwoch

Knowledge Engineering Department, Technical University of Gdansk,
80-952 Gdansk, Poland
szwoch@eti.pg.gda.pl

Abstract. In this paper an algorithm for music staves detection is presented. The algorithm bases on horizontal projections in local windows of a score image and farther processing of resulting histograms and their connections. Experiments carried out, proved high efficiency of presented algorithm and its robustness in case of non-ideal staff lines: skew and with barrel and pincushion distortions. The algorithm allows for usage of acquisition devices alternative to scanner such as digital cameras.

1 Introduction

Optical Music Recognition (OMR) is the process of converting digitized sheets of music into an electronic form that is suitable for further processing such as editing and performing by computer. More sophisticated areas of applications are automatic accompaniment, music transposing, extracting parts for individual instruments and musicological analysis of the music. The OMR systems are also the tools used for information retrieval process that takes place in creation of music digital libraries.

The OMR systems have already been developed for almost forty years that has lead to the state of high accuracy of music recognition process. Many researchers have reported the recognition efficiency over 90% of their systems [1-8]. Also several commercial systems are available on the market. In most cases, those systems operate properly only with well scanned documents of high quality. Using a scanner to digitize flat score pages assures that the only geometric distortion presented in a scanned image could be a small slant that is easily corrected by many of presented algorithms [1-10]. Only few attempts have been made to deal with scanned 3D score sources (like book pages) with non-linear bowing near the page edges [9] or with scores digitized with other optical devices [7].

The rapid growth of digital cameras' technology in last years caused that they can be considered as alternative devices for acquiring of digital images. Their essential advantages over popular scanners are high mobility and low destruction effects over digitized documents. Unfortunately, digital cameras have also several shortcomings preventing them from broad using in this domain. The main disadvantages of digital cameras are problems with picture exposure and non-linear distortions (*barrel* and/or *pincushion*) introduced to acquired images by their optics (Fig. 1).

* This paper is sponsored by the Polish Government's research funds for the years 2005-2008 as a research project No 3 T11C 027 28.

A. Gagalowicz and W. Philips (Eds.): CAIP 2005, LNCS 3691, pp. 701 – 708, 2005.

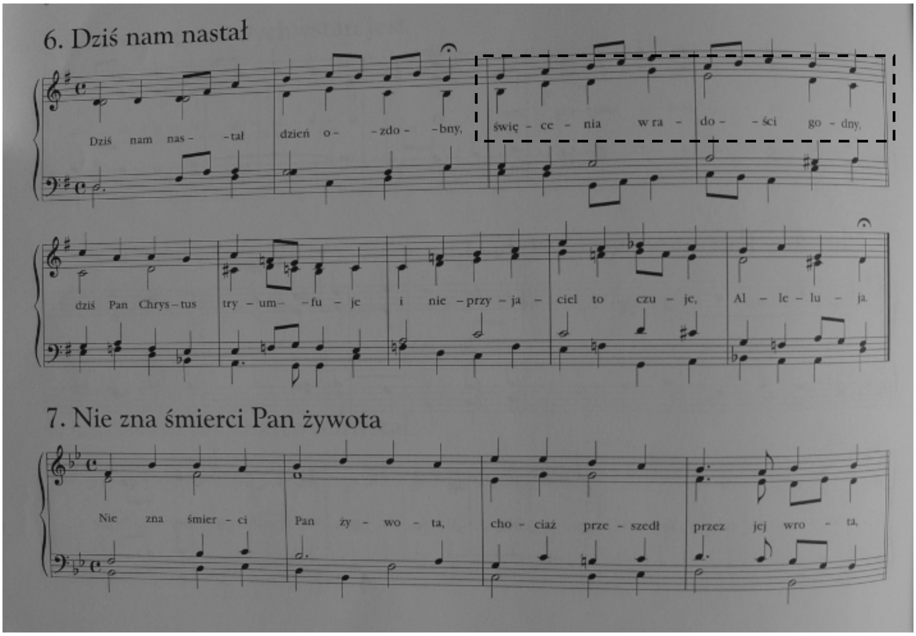

Fig. 1. Original score document taken by digital camera (5Mp) and converted to grayscale. Dotted rectangle marks the region shown as an example in Fig. 2 and 4.

In general, it is not possible to create an universal algorithm compensating geometrical distortions in images regardless of camera type and zoom settings. Fortunately, acquiring images with some regular patterns e.g. *squared* or with *staff lines*, gives a good opportunity for creation of algorithms that could detect their true shapes giving sufficient information for distortion removal. This paper concerns that aspect of *staves* detection in musical score images The presented algorithm enables to locate staff-lines with high precision giving all needed information for further image processing and recognition. The algorithm is fast and efficient for different kinds of distortions in analyzed score images.

2 Staves Location by Matching Local Histograms

The correct detection and processing of staves is fundamental to OMR process. The staff lines create a two dimensional coordinate system for interpretation of other musical symbols. The theoretically equal distance between any two subsequent lines within the same stave (and on the same page) gives the basic measure unit (called *Dist* in this paper) for other musical symbols within the same score. Finally, due to its regular linear shape, staff lines are also ideal determinant for any geometric distortions of digitized scores. Though, in most cases staff lines are removed from musical document's image, their positions are remembered for further usage during the recognition and interpretation processes.

Staves localization is not an easy task because of frequent disturbances and distortions of staff lines. Disturbances may be caused by low quality of original documents and also by other musical symbols coincident with determined lines. Geometric distortions of images are usually result of optical faults or not flat scanned surface (like books). Numerous tests proved that staff lines in digitized musical documents not always are parallel, horizontal, equidistant, of the same width or even straight!

In this section, the staves detection algorithm is presented. This algorithm bases on horizontal projections in narrow vertical strips and on farther analysis of created histograms. Many staves detection algorithms analyze *original image* searching for five equally spaced and sized run-lenghts of black pixels in [6,7,10] or straight lines using the Hough transform or mathematical morphology [4]. Unlike them, the presented algorithm operates on *histograms* what results in its low computational complexity. Additionally processing histograms of *local* projections gives it greater flexibility comparing with algorithm operating on *global* projections [1,2,5]. This flexibility allows to efficiently detect skew, bowed and spherically distorted staves.

Finally, the detected staff lines are approximated as piece-linear lines or by polynomial curves. These exact approximations allow for optional image unwarping and farther staves removal from analyzed image.

2.1 Local Horizontal Projections and Processing of Resulting Histograms

In the first stage, the analyzed image is vertically sliced into the narrow vertical strips VS_i of width equal to $2{\cdot}Dist$, numerated from 1 to N (Fig.2a). For each vertical strip VS_i, the horizontal projection's histogram H_i of black pixels is counted (Fig.2b). In the next step all histograms H_i are preprocessed in the following way:

1. Clearing histograms entries containing small values (region A in Fig. 2b-c).
2. Successive breaking of histogram columns that are wider then $1{\cdot}Dist$. Columns are broken near the local minimum values (region B in Fig. 2b-c).
3. Locating of distinct local maximum values (*peaks*) in histograms. Neighboring columns are cleared (region C in Fig. 2b-c).
4. All histograms are filtered in order to find five consecutive peaks within the distant of about $1{\cdot}Dist$. Unlike other stave filters (e.g. [6]), in this algorithm a greater level of tolerance is assumed allowing to detect lines lying in an average distance of $Dist\pm1$ pixel with variation of ±2 pixels for particular lines. After this stage all histograms contain mainly peaks representing pieces of proper staff lines (Fig. 2d).

2.2 Creating of Connection Arrays and Staves Detection

The next stage of the algorithm is creation of connections' arrays between histogram peaks HP_i^k and HP_{i+1}^l in every successive pair of histograms H_i and H_{i+1}, where k and l are vertical coordinates of histograms entries HP_i^k and HP_{i+1}^l (Fig.3). In each pair of histograms H_{i-1} and H_i, for each connection HP_{i-1}^j and HP_i^k, the linear approximation of predicted vertical position y_{pred} of the peak HP_{i+1}^{l} in the next histogram H_{i+1} is determined. Two histograms peaks HP_{i}^{k} and HP_{i+1}^{l} are then connected if their vertical distance $\left| l - y_{pred} \right| \le 0.25 \cdot Dist$ (gray region (a) in Fig. 3). If no histogram peak is found in this range, y_{pred} coordination is replaced by the last vertical coordinate, i.e. k value (range b) in Fig. 3).

Fig. 2. Stages of histograms completing: part of original image (Fig. 1) sliced into vertical strips (a), local histograms H_i of horizontal projections (b), local histograms after pre-processing (c) and after final filtering (d)

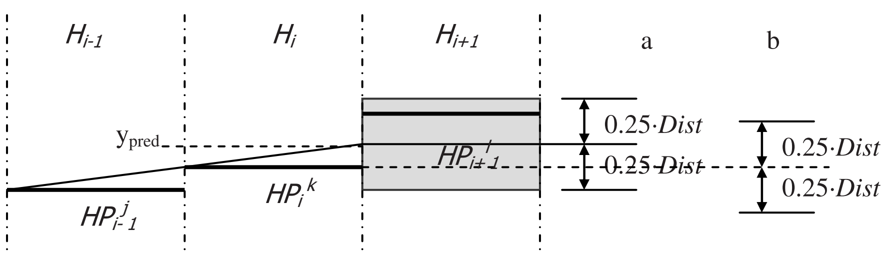

Fig. 3. Searching for connection between histograms peaks HP_{i}^{k} and HP_{i+1}^{l} using predicted vertical coordinate y_{pred} (a) or the last coordinate k (b) as a center for vertical range

a)

b)

c)

Fig. 4. Stages of stave processing: determined connections (a), final staff lines (b) and image after staves removing

As a result, an initial piece-linear approximation of staff lines (and some other symbols) is achieved (Fig. 4 a). In the next step the array of connections is processed in order to remove redundant parallel connections and to link broken connections due to locally lower quality of image or some errors in histograms processing. Further processing covers removing of short lines, gathering staff lines into staves and removing of other lines. The final result is presented in Fig. 4b. Recognized staff lines can be farther approximated by polynomial curves. Staves removing algorithm uses line parameters to efficiently remove staff lines from the image (Fig. 4c).

3 Evaluation

To validate the results of stave detection by presented algorithm 30 full-page scores had been taken by digital camera[1]. The resulting 200DPI resolution of acquired images is sufficient[2] for OMR processing [8]. The main drawbacks are serious irregular barrel distortions near the corners of the images (Fig. 1). These distortions are well visualized as curvature of detected staff lines (Fig. 5).

The described algorithm was implemented in the *ScoreExplorer* OMR system [8]. All experiments carried out proved very high efficiency of staves detection algorithm (100%). Each staff line was properly localized and its run was properly routed that was confirmed by accurate staff lines removing and farther interpretation of

[1] For all tests the 5Mp Sony FX-717 camera was used, in full automatic exposure mode, with no flash light.

[2] A good measure of score image's quality is an average staff line width. For all tested images it was 3 pixels per line.

Fig. 5. Staff lines extracted from score image presented in Fig.1

recognized musical symbols (e.g. pitch of notes). The only drawback of the algorithm of stave detection is a problem with the ends of staff lines, which are too easily extended beyond staves limits, e.g. over an accolade. In fact this problem is solved later on in OMR process when vertical elements, such as bars, are located.

The algorithm based on local projections is characterized by low computational complexity due to processing rather histograms than pixels. The average staves detection time on 4.4Mp A4 image is about 0.3s including partial visualization of detection process[3]. The processing time is short enough not to influence the whole recognition process. The algorithm could obviously be sped up by several improvements including recoding it in assembler and using every second or even every third vertical strip. Some savings may also be achieved by setting the strips width to 8 or its even multiple (16, 32, ..).

Experiments with *ScoreExplorer* proved that, despite good staves localization, the overall system's recognition efficiency decreased from 95% (for scanned images) to about 80% (for distorted images). The obvious reason is lack of tolerance for image deformations in system's recognition algorithms. The detected staves (Fig.5) could be used to eliminate distortions present in score images possibly increasing the recognition rate. Unfortunately, at the moment only simple vertical-unwarping algorithm is implemented in the *ScoreExplorer*. It corrects only vertical positions of musical symbols whereas slight slant of vertical primitives still remains (Fig. 6).

Detected staff lines can be represented as piece-linear lines or as polynomial curves. The first representation is very exact but consumes a bit more memory for all vertical coordinates of nodes (about 70 for each staff line in tested images). Experiments carried out for *scanned* scores proved that quadratic polynomial approximation of lines is sufficient but it is not so in the case of images *acquired by digital camera*. In fact irregular distortions cause that it is impossible to exactly approximate staff

[3] All experiments were carried out on a PC with Athlon 2.2 GHz processor.

lines with cubic or greater degree polynomial curves. The problem usually arises near the ends of lines which curvature is usually stronger then in the middle parts. Fortunately, in most cases the approximations errors are perceptible at the very ends of lines and do not affect recognition of other musical symbols (Fig.7).

Fig. 6. The result of simple unwarping of the example image from Fig.1

a)

b)

Fig. 7. An example of staff lines' approximation errors: a) part of original image with approximating cubic curves and b) staff lines removing errors (circled)

4 Summary

Incessant development of digital optical devices creates new possibilities of images acquiring for digital documents processing. These high-resolution devices cover not only digital cameras but in near future also Internet cameras, video cameras, handhelds and even phones. Though, the quality of their optics is generally not sufficient for pattern recognition it is possible to use them in some particular domains such as OMR.

The algorithm presented in this paper is fast and efficient for images scanned and taken by digital camera. It combines low computational complexity of histograms processing with flexibility of local methods allowing detection of straight, bowed and spherically distorted staff lines. By exactly determining the shapes of staff lines it is possible to efficiently remove them from score images enabling farther OMR processing and also to position on staves all recognized musical symbol. The experiments carried out proved that staff lines' approximation by polynomial curves is not sufficient in case of non-linear geometrical distortions of images. In that cases piece-linear approximation suits better.

Though, exact staves localization is a precondition for efficient recognition of musical documents it is not the sufficient condition in case of serious geometric distortions of source images. In that case, advanced de-bowing algorithms have to be used [9] or recognition algorithms have to be modified to take into account various deformations of recognized symbols. Developing of such algorithms will make possible to effectively use digital cameras in Optical Music Recognition.

References

1. Fujinaga I: Optical music recognition using projections, Master's thesis, McGill University, Montreal, CA, 1988.
2. Carter N.P., Bacon R.A.: Automatic Recognition of Printed Music, Structured Document Analysis, Springer-Verlag 1992, pp. 456-476.
3. Kato H., Inokuchi S.: A Recognition System for Printed Piano Music Using Musical Knowledge and Constraints, Structured Document Analysis, Springer-Verlag 1992.
4. Modayur B.R., et al.: MUSER: A prototype musical score recognition system using mathematical morphology, Machine vision and Applications, Springer-Verlag, 1993.
5. Roth M.: An Approach To Recognition Of Printed Music, Swiss Federal Institute of Technology Zurich January, 1994, Internet.
6. Reed K.T., Parker J.R.: Automatic Computer Recognition of Printed Music; Proceedings of ICPR'96, 803-807, 1996.
7. Matsushima T.: Automated high-speed recognition of printed music (WABOT-2 vision system). Proc. of the 1985 I.C. on Advances Robotics, JIRA, Tokyo, 1985.
8. Szwoch M., Meus G., Tutkaj P.: Score Explorer: A Musical Score Recognition System, In Proc. of GKPO, Warsaw 2000.
9. Wijaya K., Bainbridge D.: Staff Line Restoration, Image Processing and its Applications, Conference Publication No. 465, IEE 1999.
10. I. Leplumey, J. Camillerapp, G. Lorette: A robust detector for music staves, in Proc. of the ICDAR, pp. 902-905, Tsukuba Science City, Japan, 1993.