PLAYING CYLINDERS OF MECHANICAL ORGANS WITH AN OPTICAL READER

Laurent DAUDET, Rémi MIGNOT

Laboratoire d'Acoustique Musicale
Université Pierre et Marie Curie (Paris 6)
11 rue de Loumel, 75015 Paris, France
daudet@lam.jussieu.fr

Olivier ROMAIN

Laboratoire d'Instrumentation et des Systèmes
d'Ile de France - Equipe SYEL - B.C. 252
4, place Jussieu, 75252 Paris cedex 05, France
Olivier.Romain@lis.jussieu.fr

ABSTRACT

This study presents an experimental setup designed to read, by means of optical techniques, the music inscribed on automatic organ cylinders. We describe the acquisition principle based on images taken by a CCD linear camera, and the various digital signal processing techniques employed to retrieve the music from the images. Preliminary results show that this original method is a relevant choice, since on our test cylinder about 90 % of the notes are correctly identified, with only 14 % of false alarms. However, for realistic estimates of the actual music, some improvements are still necessary, both in the experimental setup and in the way individual note positions are converted into music.

1. INTRODUCTION: MECHANICAL ORGANS AND CYLINDERS

During most of the 18th and 19th century, mechanical organs were widespread across Europe. Actually, these devices, that could play predetermined pieces without the need of a trained musician, were in use in many churches that could not afford a permanent organist. The most employed technique for the notation of the music was to use wood cylinders, on which metallic pins of different widths mark individual notes (see fig. 1). When playing, the cylinder is rotated slowly around its axis, and the pins trigger the opening of valves, thus blowing air into the organ tubes. Because of geography, but also for "marketing" reasons, all of the many manufacturers who designed such organs [1] have used different sizes for the cylinders, different separation between the pins, and different musical scales (all the notes in the chromatic scale are generally not present).

Nowadays, public and private collections have a large number of such cylinders, while only few organs are still in playing condition. However, there is a growing interest from musicologists towards these cylinders, since they offer a privileged, and sometimes unique, testimony of the music of that time. It is very likely that some compositions exist nowadays only in that form, when the corresponding musical scores may have disappeared. Moreover, this is of utmost interest for the study of the interpretation, since the notation on cylinders accurately reproduces subtle details such as ornementation, whose exact way of playing is not written on musical scores, and can only be guessed through interpretation treaties. In other words, such cylinders are at some intermediate level between the musical score and a proper recording of the music.

In this paper, we present the preliminary stage of a study, aimed towards the design of an optical reader for such cylinders.

Figure 1: *Detail of a typical mechanical organ cylinder.*

A linear digital camera does the acquisition, and is followed by image processing techniques in order to retrieve the exact localization of the pins, and therefore the underlying music. Using optical devices has many advantages : firstly, it is contact-less, and therefore suitable for museum artifacts in very bad condition, which could be further damaged by mechanical sensors, or which are too odd-shaped for them. Secondly, it can be very easily adapted to different sizes of cylinders, since it is only a matter of proper focusing. Thirdly, it requires much less hardware than a mechanical reader, most of the work being done in software.

However, designing such a device is not without technical difficulties, since we aim at reading the cylinders in a way they were not designed for : as we shall see, even with state-of-the-art components, the signal-to-noise ratio in our images is sometimes poor. This may partly explain why this technique had not received more attention earlier.

The paper is organized as follows. In section 2, we give some details on the optical acquisition principle. Section 3 is devoted to the presentation of the image digital processing techniques employed to retrieve the written music. Finally, section 4 presents some preliminary results, and concludes on future improvements.

2. ARCHITECTURE AND PRINCIPLE OF THE ACQUISITION

2.1. Acquisition principle

The principle of the acquisition is to obtain all the information present on the surface of the Mechanical Organ's Cylinder (MOC) on a single image. The best way to achieve this specification is to acquire a panoramic image of the MOC. The panoramic image will be a planar representation of the unfolded surface of the MOC.

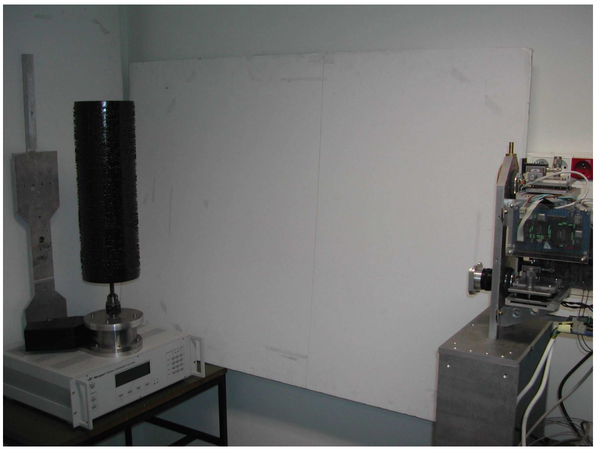

Figure 2: *Picture of the experimental setup. On the left: the cylinder on the step-by-step motor and its command box, on the right : the camera with its controller interface.*

Typical sizes for MOCs are : diameter 8 to 40 cm, length 30 to 120 cm. The pins can be as small as .5 mm wide for the shortest notes, with a precision in their position of the order of .1 mm.

2.2. Architecture of the sensor

To carry out the acquisition of a panoramic image, the MOC is fixed on a step-by-step motor in front of a linear CCD camera. Moreover, the axis of rotation of the step-by-step motor passes through the axis of revolution of the MOC and the plane of sight of the linear camera revolution includes the revolution axis of the MOC. These importants characteristics gives panoramic image without distorsion.

At each new angular position of the MOC, the image taken by the linear camera corresponds to the intersection of the plan of sight of the camera with the MOC. So, by making a continuous rotation on 360 degrees, a panoramic image is acquired. The minimum step reachable by the motor is .0001 degree.

The camera is built with a linear 1024 pixels CCD (Thomson 7802A). The photo-sensible surface of each pixel is 13x13μm. The different CCD control signals are generated by an image processing circuit (LM9800 from National Semiconductor). This circuit converts analog video signal to 8 bits digital data. Then, the digital data is stored in the FIFO RAM, before sending through an optical fiber to a dedicated ISA PC card, via a full custom serial protocol detecting transmission errors. All the others control signals are generated by an Altera FPGA. PC software carries out the data acquisition and the display of the images. A special care for reducing the camera card noise is taken all over the electronics chain. The linear camera uses a 12.5mm focal length lens. A specially designed mechanical architecture secures the alignment of the linear CCD. Three-axis micrometer screws allow a precise verticality adjustment of the linear sensor and focus.

2.3. Experimental setup

Before the acquisition of the panoramic image of the MOC, a preliminary stage of calibration is performed. It consists on one hand in making the focus of camera by using a classical chart, and on the second hand, in making sure that the MOC is under a constant illumination. For that, we have used light sources whose spectral characteristics are close to day light at 5400° K.

Under these considerations and with regards the technical specifications, panoramic image with a great resolution can be taken. Experimentally, with a typical angular step of .1 degree, we obtain sizes of 3600 by 1024 pixels. The linear camera is placed at a distance of approximatively one meter of the MOC. Due to the monochromatic property of the camera, only black and white panoramic images can be taken. Color panoramic image is possible by adding some special filters [2].

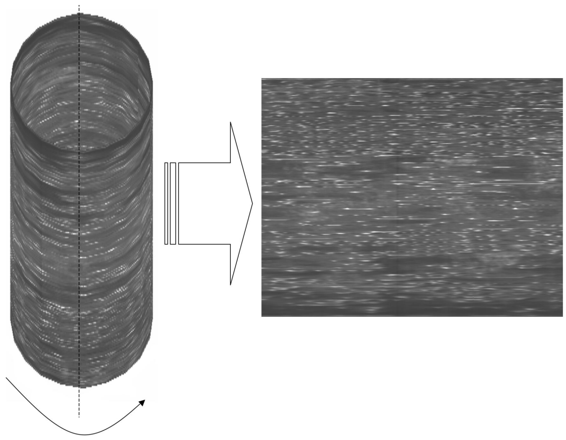

Figure 3: *The resulting image is obtained by unfolding the surface of the organ cylinder.*

3. FROM IMAGES TO MIDI SCORES

Once the images have been obtained, they have to be digitally processed in order to retrieve the underlying music. For its simplicity and because of its widespread use, we have chosen MIDI as our target format.

3.1. Pre-processing

Pre-processing the images is essential to obtain reasonably good results in the estimation process, mainly to remove most of the artifacts due to the irregularity in lightning and varnish at the surface of the MOC. Best results have been obtained by using an adaptive filter, obtained by training on a limited portion of the image where we know the ground-truth desired output.

In order to get this filter, the process is as follows (see fig. 4). With a 2D high-resolution color digital camera, we take the picture of a small portion of the cylinder. The size of this image must be carefully chosen, since it has to be small enough so as to assume that the surface of the cylinder is approximately flat, but also it has to contain a sufficient number of pins (typically a hundred) for the training algorithm to work properly. Taking this image as reference (image K), we draw by hand an idealized "ground-truth" image J, with constant-width monochrome lines at the exact position of the pins. The optimal filter h^* would be the one that transforms K into J :

$$J = Kh^* \tag{1}$$

The least-squares estimate $\tilde{h}$ of h^* is obtained by taking the pseudo-inverse of K :

$$\tilde{h} = (K'K)^{-1}K'J \tag{2}$$

where K' denotes the complex transposed of K.

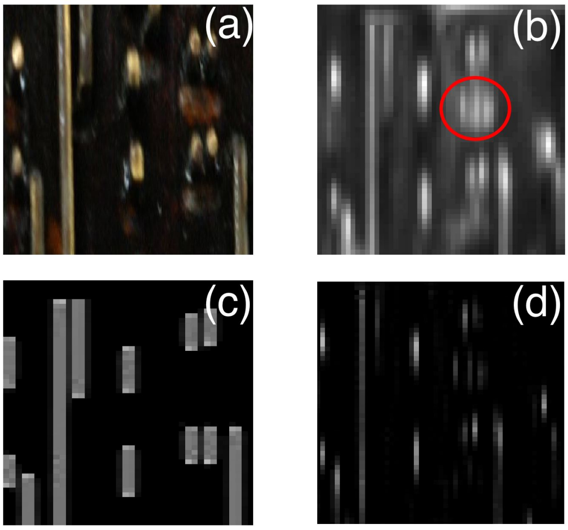

Figure 4: *Pre-processing the images with an adapted filter.* ***(a)*** *Small portion of the image shot with a 2D high-resolution color digital camera.* ***(b)*** *Same portion taken with the 1D camera. Notice (red circle) that the spot created by irregularities in varnish appears here like (false positive) pins.* ***(c)*** *Ideal image created by reproducing by hand the position of the pins of image (a).* ***(d)*** *Result after processing image (b) with the optimal filter.*

This approach usually gives satisfactory results, as most of the artifacts disappear. However, it is still unable to distinguish between pins and some varnish irregularities, therefore prone to false positives in the detection stage. Also, drawing the idealized image is a very tedious task even on a small image. For on-site experiments, this approach is not applicable as such, and more classical denoising / contour detection approaches may have to be employed.

3.2. Estimation of the position of the individual tracks

After preprocessing, the next task is to identify the position of the individual tracks of the different notes, for a given piece of music. Actually, on a given MOC as much as 10 different pieces are present at the same time, with an interleaved positioning of the notes (cf fig. 5). This configuration allows an easy switch between two pieces (one only has to mode the cylinder sideways by a couple of mm), together with keeping the distance between two notes of the same piece relatively large. A trained observer is usually able to identify the number of notes in the scale, and the number of pieces written on a cylinder : usually the distance between tracks of the same note is lower than the distance between adjacent notes. In other words, a cylinder containing P pieces of music on a scale of N notes will exhibit N groups of P tracks each. On our test cylinder, we had $N = 24$ groups of $P = 10$ tracks each, therefore a total number of 240 tracks.

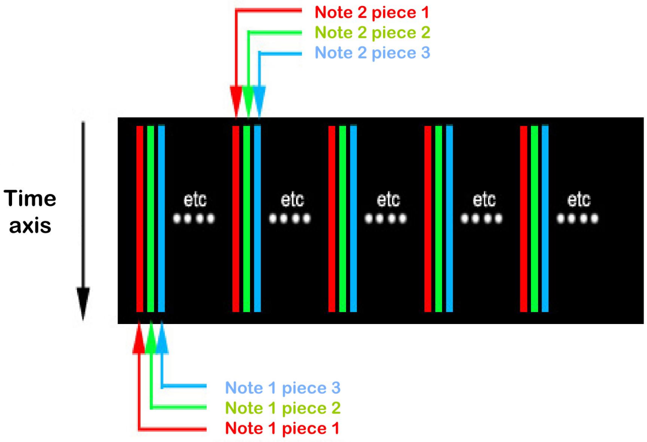

Figure 5: *Different pieces of music are interleaved to keep constant the spacing between notes.*

The easiest approach is to select as relevant columns of the images (*i.e.* tracks for individual notes) the ones where the "energy" (sum of the squares of the pixel value) is higher than a given threshold. Unfortunately, this method fails for two reasons : first, artefacts bring energy into columns that do not correspond to tracks (false positive); second, some notes of the scale may not be used in one or more pieces, in which case the corresponding tracks are not detected (false negatives).

Here, we have chosen an optimization technique to improve the above results. Let us call our unknown parameters as : d the distance between adjacent tracks within a group, e the distance between adjacent tracks that separate two groups, and b the position of the first track.

The position $n_{b,e,d}(i,j)$ of the tracks given by the model is

$$n_{b,e,d}(i,j) = b + d*(i-1) + ((P-1)d + e)*(j-1) \tag{3}$$

for the j-th note (j = 1 ...N) of the i-th piece (i = 1 ...P). The optimal triplet (b, e, d) is the one that minimizes on average the distance between the position of the tracks given by the model, and the position of the closest tracks obtained experimentally by the energy method described above. Due to the non-continuous property of the above cost function, it has many minima and one has to use a "brute force" method of scanning the whole space of parameters within a reasonable range. Optional improvements such as consistency constraints on the displacement vectors (distance between the model and the observed track) allows further refinements of the above estimates.

Finally, it is now possible to use the estimates on (b, e, d) to gather the columns of the image that correspond to only one musical piece.

3.3. Adaptive thresholding

On a given column, one has to threshold the pixel values : pixels higher than the threshold are considered as part of a note pin.

However, differences in lightning conditions imply that the same threshold cannot be used on all columns. Here, we use a different threshold on each group of P tracks, chosen adaptively using Otsu's method [3] that divides the pixels into two classes using first and second-order moments.

On the binary images, the position and width of each of the pins can now be retrieved. On our test example, the results were as follows :

Number of pins	163
Good detection	148
False alarms	23
Non detection	15
Double (/triple) notes	13

which, given the relatively bad condition of the cylinder, can be considered as acceptable.

3.4. Conversion to MIDI format

The last step is to convert the position of the notes into MIDI format, where the corresponding music can be played and edited for a manual post-processing on commercial software. We have used the MIDI Toolbox [4] for Matlab, available under GNU public licence. This toolbox allows the conversion of piano-roll type of images to MIDI notation. Here, an unknown is the scale that is used : in general this is neither a chromatic scale nor a scale given by one (or more) tonality, and it may well be an intermediate between the two : chromatic for the higher pitch range, but tonal in the lower range. Other unknowns are the sense and speed of rotations, but can be guessed will little prior knowledge. For instance, the speed of rotation can be roughly estimated if we assume the total duration of each piece to be of the order of 1 minute.

4. PRELIMINARY RESULTS AND FUTURE IMPROVEMENTS

Although the note recognition process gives acceptable results, the resulting soundfiles, played on a MIDI synthesizer, cannot be recognized as musical pieces. Some patterns are identified, such as chords, or ornementation, but the "big picture" is missing. One one hand, we have many unknowns, mostly about the scale that is used, but also, on a more basic level, on the direction and speed of rotation of the cylinder. On the other hand, we have some transcription errors, and this prevents listeners (such as the authors of this article) that are not organ musicologists to have relevant guesses about the above unknowns. Therefore, more work is needed in a few directions :

- the acquisition setup has to be improved. Indeed, on our test cylinder we wanted to retrieve 240 tracks of music with a 1024 pixels linear camera, and clearly this resolution is too tight. With this configuration, a given track can appear on only one pixel or be shared by two pixels, and this makes the whole system very shift-dependent. Ideally, we would like to have a larger number of pixels per track, as well at least one pixel between two tracks. Future plans include the use of a higher-resolution CCD camera, for instance with a line of 8192 pixels. Also, the varnish is a large source of transcription errors, and it should be possible to use a source of light with only wavelengths that are not absorbed by the varnish.
- with an increased resolution, the image processing technique will have to be modified. Even now, the storage size of each image of the whole cylinder is quite large, and ideally one would like to design an algorithm that processes the images on-the-fly during the acquisition, without having to store the unprocessed images.
- interaction with musicologists and specialist museums is strongly needed. In fact, without their knowledge of the intruments, the scales most commonly used in such organs, and the type of music played on them, it is virtually impossible to guess all the unknown mentioned above. Hopefully, with such knowledge, we would be able to obtain results that can be compared to real musical pieces; and this is ultimately what our technique must be evaluated on.

However, this work shows a proof of concept that optical reading is a viable method for retrieving the music written on mechanical organ's cylinders, although its practical implementation requires state-of-the-art hardware and software techniques. Finally, this method should not be seen as an alternative to the restoration of historical mechanical organs ; on the contrary it may help specialists to determine which of the organs should undergo an expensive restoration on the grounds of musicologist interest.

5. ACKNOWLEDGEMENTS

We would like to acknowledge Noémie Hassan and Vincent Herrgott for their contribution to this study.

6. REFERENCES

[1] Dom Bedos de Celles, *L'Art du Facteur d'Orgue, Qatrième partie*, Bärenreiter Verlag, 1966, fac simile from the original 1778 edition.

[2] O. Romain, T. Ea, C. Gastaud, and P. Garda, "A multi-spectral sensor dedicated to 3D spherical reconstruction," in *Proc. of IEEE Int. Conf. Image Processing (ICIP'01)*, Thessaloniki, Greece, Oct. 2001.

[3] N. Otsu, "A threshold selection method from grey scale histogram," *IEEE Trans. on Syst. Man and Cyber.*, vol. 1, pp. 62–66, 1979.

[4] T. Eerola and P. Toiviainen, *MIDI Toolbox: MATLAB Tools for Music Research, available at* `http://www.jyu.fi/musica/miditoolbox`, University of Jyväskylä, Finland, 2004.