

A Proposed Optical Music Mark Recognition System

Ismael Jannoud

Damascus University, Damascus, Syria and Al-zaytoonah University, Amman, Jordan

Abstract: The optical music symbol recognition system still has many challenges. Most of issues focused on the recognition of the individual marks in the input concert documents. In this study a new and efficient model for optical music recognition (OMR) was proposed. The music marks was recognized and played by the computer OMR system automatically. A stored file of music documents has been treated sequentially as inputs to the off-line recognition system. A new algorithm for staff-lines removing was proposed also. This new technique is applied on the thinned image (document) without affecting the marks. Features have been extracted from isolated marks and stored as a training data. A wave files has been stored for the same marks played by a guitarist as an example. Minimum distance classifier is used as a recognition approach. Proposed model improves the music information technology (MIT) and develops music academies courses, music distance learning. That leads to enhance the business process of developing music and music scientists.

Key words: Optical music recognition (OMR), music document, music information technology (MIT)

INTRODUCTION

Optical music symbol recognition system is a stream of the recognition science. Refers to the machine recognition system of page context of the printed music notes, these pages are optically scanned and stored as gray image files. The model is partitioned into two major steps. The first is the *learning* stage which gives the training data and the second is the *test* stage which depends on one of the classification approaches to test the unknown input data to the system^[1]. Figure1 display the first stage in detail and the last two steps refers to the second stage. Training data is obtained for a large number of music marks. The concert document passed through the system and some pre-processing steps applied on it (binarization, de-noising, slop correction, thinning and staff-lines removing). Some of these steps are common for many gray images applications like optical character recognition. The others are special for the OMR model. This study focused on the final post-processing step as a challenge target of this model. This step is done by make the computer not only recognize the mark, but trying to play it directly. The model takes one music machine; the guitar; as an example in the first version of the OMR model. The steps discussed in brief in this study. A new algorithm for the staff-lines removing is proposed. Results from some processing steps were listed, also.

The proposed Model aims to enhance the business process of developing music academies. All music academies needs to manage and process the music courses, developing their internet sites, restoring music, music source classification and other applications in MIT^[2,3].

Data acquisition: The printed music documents is scanned and stored in bitmap files of gray image type with low resolution. An adaptive threshold is used for the binarization step to converting image from gray into black and white. The *dominant gray level* value in the image is determined. The threshold then is the mid point between the *dominant value* and the maximum gray value in the image^[4]. A black and white image is obtained (all the back ground pixels set to zeros “0” and the information set to ones “1”). The next step is the noise reduction (de-noising).

De-noising: Noise may form as one pixel of value “1” without connectivity. It may caused by the data acquisition system. A smoothed algorithm may be applied on the original image to eliminate these individual ones “1”.

Templates of 3x3 pixels is used to examine each pixel (Fig. 2). Image is searched, if any pixel has value “1” and match with its neighbors one of the four templates of Fig. 4; X’s: refers to don’t care values (“0” or “1”); this pixel should be changes to zero.

Thinning: This step is applied on the image after slop normalization to get horizontal staff-lines. Thinning algorithm keep only the skeleton of each object in the image. The algorithm listed in^[5]. Figure 3a, b shows an image sample before and after thinning.

Sub-division and component labeling: In order to reduce processing time and keep complexity of the code low, only one staff at a time is processed once the staff lines are located on page. Lines of the concert document have been isolated before the removing the

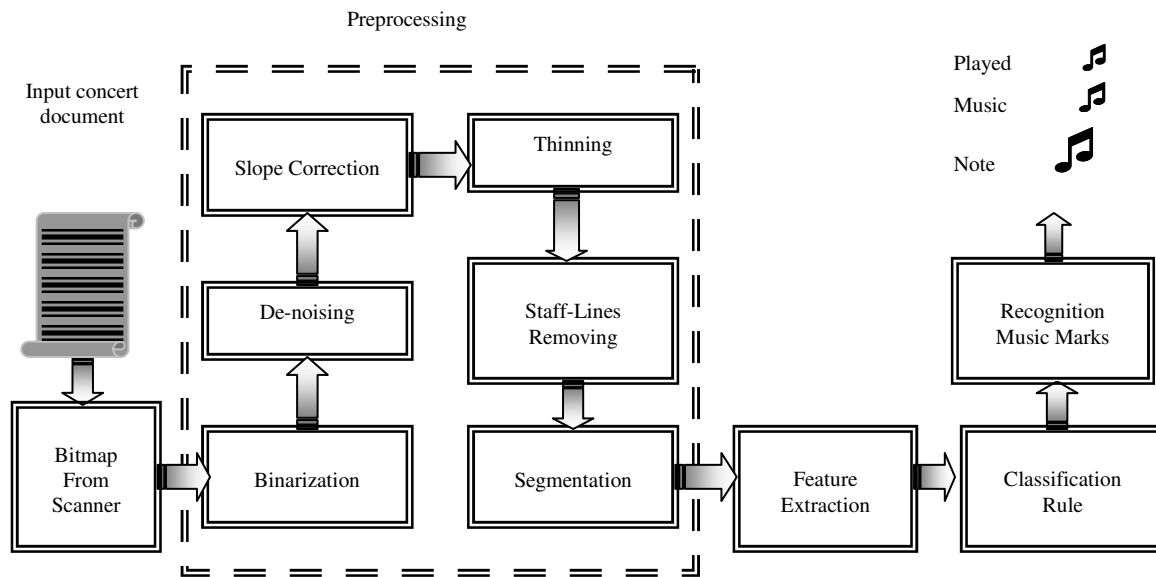


Fig. 1: Proposed OMR model

0	0	0	x	0	0	x	x	x	0	0	x
0	1	0	x	1	0	0	1	0	0	1	x
x	x	x	x	0	0	0	0	0	0	0	x

Fig. 2: Smoothing templates



Fig. 3a: Notes before the thinning

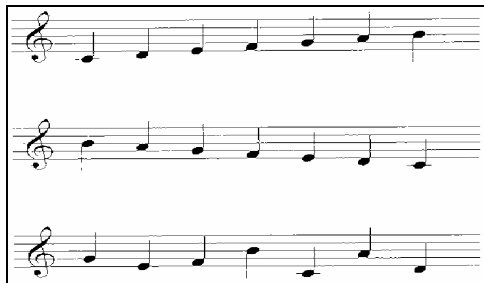


Fig. 3b: Notes after the thinning

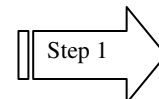
staff-lines of each. A fixed space above and below the staff is included in the region of interest; the total height of the region is three times the staff width. This allows for the recognition of up to four ledger lines, if the stem direction is conventional.

Staff-lines removing: The proposed technique for staff-lines elimination is more efficient in comparison with the traditional technique. Most of the used techniques^[1,4,6,7] make a distortion in the marks of the concert document. Thus a restoration step for the lost information need to applied on the image. The proposed technique has been applied on the thinned image. The algorithm based on the same assumption (all the information in the processed image are ones “1” and the back ground pixels are zeros “0”). The two steps of algorithm are (in brief) are:

Step 1: Search on the thinned image for the template (Fig. 4a). Pixels of the second line related to the thinned staff-line. If this template matched anywhere in the sub-image (single line of document), change the center pixel which have value “1” into “0”. Figure 4c shows the sub-image after this step.

Step 2: Search sub-image again. If a templates of Fig. 5a, b matched in the sub-image, change the ones in the second line into zeros. It means each pixel has value “1” and the pixels above and under are “0”, change this pixel into “0” (Fig. 5c).

0	0	0
1	1	1
0	0	0



0	0	0
1	0	1
0	0	0

Fig. 4a: Template

Fig.4b: Template after step 1

Component labeling: The following step is the mark isolation. Any traditional technique may applied on the sub-image (line concert document). If the document include a very cursive printed marks, a coloring

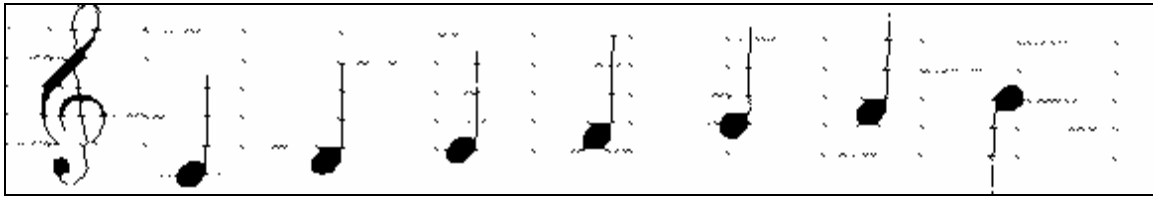


Fig. 4c: Staff-line removing after applying step 1

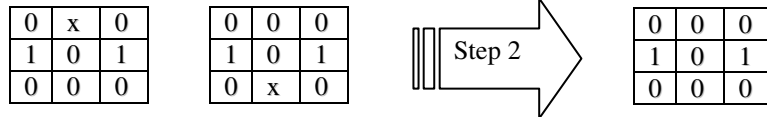


Fig. 5a and b: Templates before step 2

Fig. 5c: Template after step

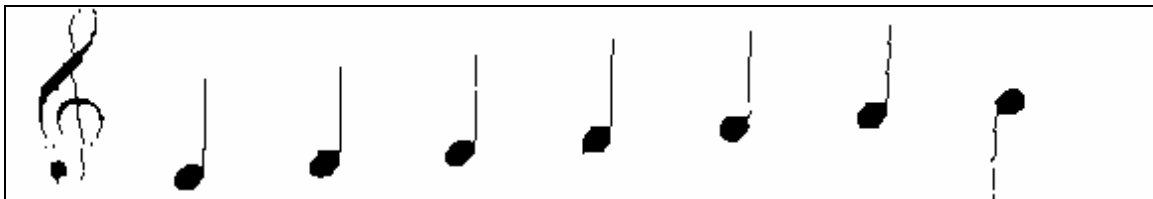


Fig. 5d: Staff-line removing after applying step 2

(labeling) algorithm should be used to label each connected components (pixels) related to the same mark^[8].

A border around each sub-image is achieved and each mark has been isolated in a sub-matrix. All marks matrices are in the same size.

Features extraction: The marks are (*Do, Re, Me, Fa, So, La, Ce*) treated in separated two groups. The first contains the all without *Ce* and the second contains *Ce* mark only. First of all for the first group it is required to distinguish between the *BLANCH music marks* (symbols with fill bottoms) and the *NWAR music marks* (symbols with no fill bottoms). Thus each matrix of individual mark has been bordered exactly around the mark area. The area then partitioned horizontally into two equally parts. The first two features are the area, i.e. the number of one's of each part (to decide whether the mark is *BLANCH* or *NWAR*). Then each part is partitioned again into six sub-blocks. Two features are extracted from each (the width and the area in each and the area in each sub-block). All of these features in addition to the position of the center of the original music mark are normalized according to the same feature from the whole mark. The *Sol Key* (&) mark shown at the beginning of each staff-line has the largest area and width (after isolation the individual marks). Features are extracted for a large number of samples and stored as a training data.

Recognition and training data: The recognition algorithm works with the set of extracted features. A minimum distance classifier has been used. The classified mark is labeled and added to the list of

recognized components. For classification as a certain element, once classified, the component is not looked at any more, i.e. the classification is never changed.

Model Test and Classification: As mentioned before, we used a 15-dimensional pattern vector X which defined as:

$$X = [x_1 \ x_2 \ \dots \ x_D]^T \quad (1)$$

For classification a multi-category classifier is used. The training samples are used to estimate the mean vector $\underline{\mu}_i$ and the covariance matrix \sum_i for each class ω_i ^[9]. These are known as the *Maximum Likelihood Estimation* (MLE) of the parameters, where n_i is the number of samples for class ω_i . For each set of marks, a subset of the feature set is selected. This selection is the *Sequential Forward Selection* technique (SFS)^[10]. This technique is based on choosing the best single feature, then the best pair including the best feature and so on. Then, a subset of the feature set has been used which gives a maximum probability of mark classification. This technique is simple, efficient and produces a sub-optimal set of features, not all features. To classify an unknown mark, its feature vector \underline{X} is calculated and a generalized distance measure between this vector and each of the average vectors is evaluated, using the selected features. This distance measure, known as the *Mahalanobis* distance, is defined as:

$$DIST_i = (\underline{X} - \underline{\mu}_i)^T \sum_i^{-1} (\underline{X} - \underline{\mu}_i) \quad (2)$$

This distance is used to classify unknown mark using the following minimum-distance classification rule:

$$\text{Choose Class } \omega_i \text{ If } (DIST_i \leq DIST_j, \forall j \neq i) \quad (3)$$

Thus, the unknown music mark is assigned to the class corresponding to the closest average vector in the feature space, after feature selection.

Enhancement of the MIT: Each region has its special music. Usually this music played using special devices, also. The proposed OMR model may guess the music class by detection of the frequency of music marks and the repetition of some note segments. This target is achieved partially in this model and still under research in continues of this work.

CONCLUSION AND FUTURE WORK

The proposed OMR model has been implemented using two programming languages (C++ and MATLAB). The performance was about 93% when testing the model. A training data with related wave files has been stored for played music by a guitarist. The work is still under development to continuing process the other music machines to form a great music information data. This data can be treating by the music scientists and develop the music. Proposed model improves the music information technology (MIT) and develops music academies courses, music E-learning. That leads to enhance the business process of developing music. Another benefit is that pages of any concert may be coded using the proposed OMR model and the training data.

ACKNOWLEDGEMENTS

I wish to sincerely thank the director of the National Institute of Music, for encourage and help.

REFERENCES

1. Bainbridge, D., 1991. Preliminary experiments in musical score recognition. B.Eng Thesis. Dept. of Computer Science, Univ. Edinburgh.
2. Introduction to MIDI and Computer Music. Course in 3ed Cycle in School of Music, Indiana University.
3. Byrd, D., 1997. Independent research in music information technology. Indiana University, Spring 2003.
4. Clarke, T. and H. Baird, 1997. Optical Music Recognition Report. Sep. 4.
5. Rafael, C.G. and R.E. Woods, 1987. Digital Image Processing. Prentice Hall.
6. Couasnon, B., P. Brisset and I. Stephan, 1995. Using logic programming languages for optical music recognition. Paris, France.
7. Parker, J.R., 1997. Algorithms for Image Processing and Computer Vision. Prentice Hall.
8. Altuwaijri, M. and M. Bayoumi, 1994. Recognition of arabic characters using neural networks. ICECS, pp: 720-725. Dec. 19-22, Cairo. Egypt.
9. Devijver, P. and J. Kittler, 1982. Pattern Recognition: A Statistical Approach. Prentice Hall, Englewood Cliffs, N.J.
10. Robert, J.S., 1992. Pattern Recognition: Statistical, Structural and Neural Approaches. New York.