Optical Music Recognition System within a Large-Scale Digitization Project

G. Sayeed Choudhury†, Tim DiLauro†, Michael Droettboom‡,
Ichiro Fujinaga‡, Brian Harrington† and Karl MacMillan‡

†Digital Knowledge Center, Milton S. Eisenhower Library
Johns Hopkins University
and
‡Peabody Conservatory of Music
Johns Hopkins University

Abstract

An adaptive optical music recognition system is being developed as part of an experiment in creating a comprehensive framework of tools to manage the workflow of large-scale digitization projects. This framework will support the path from physical object and/or digitized material into a digital library repository, and offer effective tools for incorporating metadata and perusing the content of the resulting multimedia objects.

1 Introduction

The project involves digitization of the Lester S. Levy Collection of Sheet Music (Milton S. Eisenhower Library, Johns Hopkins University). Central to this project is the optical music recognition program, which converts bit-mapped images of scanned music into machine-readable formats. In Phase One, images of the music and lyrics, and color images of the covers of the Levy Collection were digitized and a database of text index records was created. Phase Two consists of converting the digitized music to computer-readable music notation format along with full-text lyrics, generating sound renditions, and creating metadata to enhance search capabilities.

2 Lester S. Levy collection

During Phase One, the researchers at the Eisenhower Library created a database of text index records, images of the music and lyrics and color images of the cover sheets from the Levy Collection. This database is available to the general public at `http://levysheetmusic.mse.jhu.edu`.

The Levy Collection consists of over 29,000 pieces of popular American music. While the Collection spans the years 1780 to 1960, its strength lies within its thorough documentation of nineteenth and early twentieth-century America through popular music. The Collection also provides a social commentary on American life and a distinctive record of their time.

Currently, the Collection can be searched in three modes. First, users can search by subject, a keyword search on the text record. Each of the pieces has been indexed for the subject of the song and/or cover image. Users may also browse the Collection by the topical arrangement of the physical collection. The physical collection's organization scheme includes 38 topics, such as the circus; dance; drinking, temperance and smoking; fraternal orders; presidents; romantic and sentimental songs; schools and colleges; and transportation. Finally, users with interest in the graphical elements can examine the Collection by focusing on the cover art.

3 Digital workflow management

The entire project is an experiment in developing a comprehensive framework of tools to manage the workflow of large-scale digitization projects. This framework will support the path from physical object and/or digitized material into a digital library repository, and offer effective tools for incorporating metadata and perusing the content of the resulting multimedia objects. The Levy Collection, with its large size and availability in digital format, is an ideal subject for development and evaluation of this proposed framework.

Phase One of the Levy Project focused on digitally scanning the music into TIFF files, converting to JPEG images and thumbnails, and then mounting the images on the Web. Online indexing was also created at the sheet music item level. An index record was created for each piece of music, which included when available: the unformatted transcription of title, statement of responsibility, first line of lyric, first line of chorus, dedication, performer, artist/engraver, publication information, plate number, and box and item number. We also introduced controlled vocabulary in the form of brief subject terms for both the content of sheet music covers and content of songs from the Library of Congress's Thesaurus for Graphic Materials. Currently the information is available as unformatted free text files that can be searched by simple keyword or phrase.

In Phase Two, an adaptive optical music recognition (AOMR) software (Fujinaga 1997) is used to convert the TIFF image of scanned sheet music into computer readable-formats, which includes GUIDO and MIDI files along with full-text of the lyrics. These digital objects will be deposited into the data repository along with the scanned sheet music (TIFF, JPEG, thumbnail, and associated metadata).

4 Adaptive optical music recognition

The AOMR software offers five important advantages over similar commercial offerings. First, it can be run in batch processing mode, an essential feature for the Levy Collection given its large number of music sheets. It is important to note that most commercial software is intended for the casual user and does not scale for a large number of objects. Second, the software is written in C and therefore is portable across platforms. Third, the software can "learn" to recognize different music symbols—an issue considering the diversity of the Levy Collection specifically, and the universe of notated music, in general. Fourth, the software is open-sourced. Finally, this software can separate full-text lyrics that can be further processed using third-party optical character recognition (OCR) software. Preliminary attempts at using the existing OMR system for OCR also show some promise.

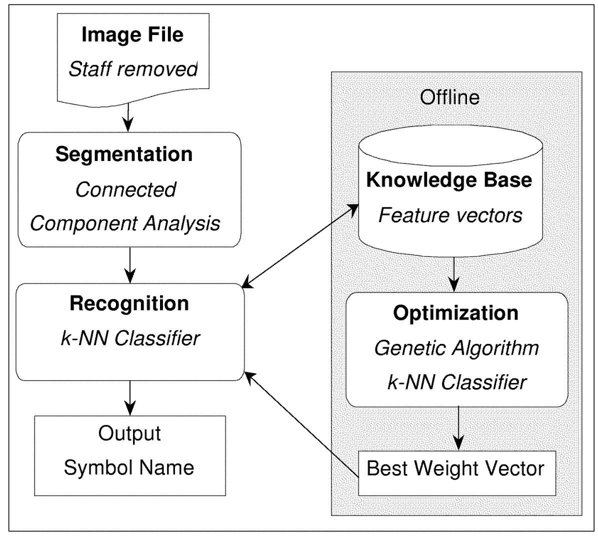

Figure 1: Overall architecture of the AOMR system

Both output formats of the AOMR, GUIDO and MIDI, represent non-proprietary file formats. GUIDO is a file format designed to allow the interchange of music notation data between and among music notation editing and publishing programs and music scanning programs (Hoos and Hamel 1997). MIDI provides low-bandwidth transmission of music over the Internet so that the users can listen to the music with their web browsers.

Recently the AOMR software was ported to the Linux environment using the GTK+ library (Pennington 1999). The overall architecture of the system is shown in Figure 1. The console-only version also compiles and runs on Microsoft Windows using the mingw32/gcc tool chain.

Using vertical run-length coding and projection analysis the staff lines are removed from the input image file. Lyrics are also removed using various heuristic rules. The music symbols are then segmented using connected-component analysis. A set of features, such as width, height, area, number of holes, and low-order central moments, is stored for each segmented graphic object and used as the basis for the adaptive recognition system based on examples.

The exemplar-based classification model is based on the idea that objects are categorized by their similarity to stored examples. The model can be implemented by the *k*-nearest-neighbor (*k*-NN) algorithm (Cover and Hart 1967), which is a classification scheme to determine the class of a given sample by its feature vector. Distances between feature vectors of an unclassified sample and previously classified samples are calculated. The class represented by the closest neighbor is then assigned to the unclassified sample. Besides its simplicity and intuitive appeal, the classifier can be easily modified, by continually adding new samples that it "encounters" into the database, to become an adaptive system (Aha 1997). In fact, "the nearest neighbor algorithm is one of the simplest learning methods known, and yet no other algorithm has been shown to outperform it consistently" (Cost and Salzberg 1993). Furthermore, the performance of the classifier can be dramatically increased by using weighted feature vectors. Finding a good set of weights, however, is extremely time-consuming, thus a genetic algorithm (Holland 1975) is used to find a solution (Wettschereck, Aha, and Mohri 1997). Note that the genetic algorithm can be run off-line without affecting the speed of the recognition process.

5 Interpretation of musical semantics

After the symbols have been classified, their musical semantics are interpreted (Droettboom 2000). While other approaches to this problem have used relatively high-level data structures such as graph grammars (Fahmy and Blostein 1998) or temporal trees (Diener 1990), our method proves that is possible to perform most operations quite conveniently using a simple, temporally-sorted list.

The first step in the interpretation phase is to connect all inter-related symbols. For instance, in order to determine the pitch of the note, it must be related to a set of staff lines, a clef and a key signature. Many rhythmic errors can also be corrected by adjusting the metric placement of notes relative to their vertical alignment with notes in other parts.

The interpretation is performed using a highly dynamic object-oriented system implemented in Python. This design makes it possible to define the semantics of new symbols at run-time, further facilitating the extensible nature of the AOMR system.

6 Current developments

An interactive graphic editor suitable to be interfaced with the AOMR program is being developed jointly with the group working on the GUIDO editor (Renz 2000). The purpose of this editor is to correct any errors generated by the AOMR so that the corrected version then can be converted to GUIDO format. The GUIDO editor was chosen because its internal data structure, GUIDO Graphic Stream (GSS), is very similar to the original Postscript output of AOMR.

To enable powerful and fast music search retrieval, we are adapting an advanced natural language search engine to locate various musical attributes, such as melodic phrase, interval, and contour. To enable this searching and provide a user-friendly navigational mechanism, Phase Two of the Levy Project will include a strong metadata component. Commonly defined as "data about data," metadata is structured representational information. The kinds of metadata important for Levy include descriptive (to enable searching, browsing and identification of items), structural (to enable the creation of an interface for optimum browsing and navigation), and administrative (to manage the digital components of the collection and aid users in identification of items).

The current index-text will be converted into structured metadata using XML (Extensible Markup Language). This metadata will be bound, along with the music notation, image, sound, and lyric data, into a complete digital sheet music object. We will use the XML markup to create indexes that will allow users to move between general keyword and precise searches. We will extract rich name information from the unstructured index-text into specific indexes such as composer, lyricist, or arranger, and possibly performer, artist, engraver, lithographer, dedicatee, and publisher. Cross-references will direct searchers to index records containing varying forms of names, including pseudonyms, transcribed from the sheet music pieces. All records will have the "authoritative" version of names. The subject terms will also receive mark-up to facilitate subject keyword searching.

To further enhance the scholarly value of the Levy Collection, a web interface will be developed for a music research toolkit, for example, Humdrum (Huron 1997). These toolkits are software tools intended to assist in music research and are suitable for use in a wide variety of computer-based musical investigations, such as motivic, stylistic, and melodic analysis and concordance studies.

We also propose to extend plans for developing automated means of mining authoritative name information and creating even richer name indexes.

7 Conclusions

The Levy Collection, with its large size and availability in digital format, is an ideal subject for development and evaluation of a comprehensive framework of tools to manage the workflow of large-scale multimedia digitization projects. The adaptive optical music recognition program will play a central role in this endeavor.

The end result should significantly increase the collection's research value to scholars, educators, writers, musicians, and the general public, taking advantage of searching both musical and textual data, viewing the digitized sheet music and covers, linking to full-text lyrics, and hearing sound files.

8 Acknowledgements

This project is funded in part by the National Science Foundation, the Institute for Museum and Library Services, and the Levy family.

References

Aha, D. W. (1997). Lazy learning. *Aritificial Intelligence Review 11*(1), 7–10.

Cost, S. and S. Salzberg (1993). A weighted nearest neighbour algorithm for learning with symbolic features. *Machine Learning 10*, 57–78.

Cover, T. and P. Hart (1967). Nearest neighbour pattern classification. *IEEE Transactions on Information Theory 13*(1), 21–7.

Diener, G. (1990). *Modeling Music Notation: A Three-Dimensional Approach*. Ph. D. thesis, Stanford University. Stan-M-69.

Droettboom, M. (2000). Design and implementation of an optical music interpreter. Technical report, Peabody Conservatory of the Johns Hopkins University. http://www.peabody.jhu.edu/ mdboom/omi/.

Fahmy, H. and D. Blostein (1998). A graph-rewriting paradigm for discrete relaxation: Application to sheet-music recognition. *International Journal of Pattern Recognition and Artificial Intelligence*.

Fujinaga, I. (1997). *Adaptive optical music recognition*. Ph. D. thesis, McGill University.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Hoos, H. H. and K. Hamel (1997). GUIDO music notation version 1.0: Specification part I, basic GUIDO. Technical Report TI 20/97, Technische Universität Darmstadt. http://www.informatik.tu-darmstadt.de/AFS/GUIDO/docu/spec1.htm.

Huron, D. (1997). Humdrum and kern: Selective feature encoding. In E. Selfridge-Field (Ed.), *Beyond MIDI: The Handbook of Musical Codes*, pp. 375–401. Cambridge, MA: MIT Press.

Pennington, H. (1999). *GTK+/Gnome Application Development*. Indianapolis, IN: New Riders.

Renz, K. (2000). Design and implementation of a platform independent guido notation engine. In *International Computer Music Conference*, pp. 469–472.

Wettschereck, D., D. W. Aha, and T. Mohri (1997). A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review 11*, 272–314.