Handling Scanned Sheet Music and Audio Recordings in Digital Music Libraries

C. Fremerey[1], D. Damm[1], M. Müller[2], F. Kurth[3], M. Clausen[1]

[1] *Bonn University, Dept. of Computer Science III, Germany, Email: {fremerey, damm, clausen}@cs.uni-bonn.de*

[2] *Saarland University and MPI Informatik, Germany, Email: meinard@mpi-inf.mpg.de*

[3] *FGAN-FKIE, Abteilung KOM, Wachtberg, Germany, Email: kurth@fgan.de*

Introduction

The last years have seen increasing efforts in building up large digital music collections. These collections typically contain various types of data ranging from audio data such as CD recordings to image data such as scanned sheet music, thus concerning both the auditorial and the visual modalities. In view of multimodal searching, navigation, and browsing applications across the various types of data, one requires powerful tools that support the process of analyzing, correlating, and annotating the available material.

In this paper, we handle two types of digital music documents: scanned sheet music and audio CD recordings. Given sheet music and audio recordings for the same pieces of music, we have representations of the same abstract musical work in two different domains. Our goal is to make extended use of the availability of such multi-domain information to close the gap between the domains in user applications. Traditionally, a user would work with either sheet music or audio data in an isolated fashion. On the contrary, in our application the user can interact with a piece of music using both sheet music and audio representations at the same time. Figure 1 shows screenshots of our user interface for displaying and interacting with the two representations. The merging of domains clears the way for a more powerful handling of the content through cross-domain tasks involving matching, retrieval and synchronization across the domains.

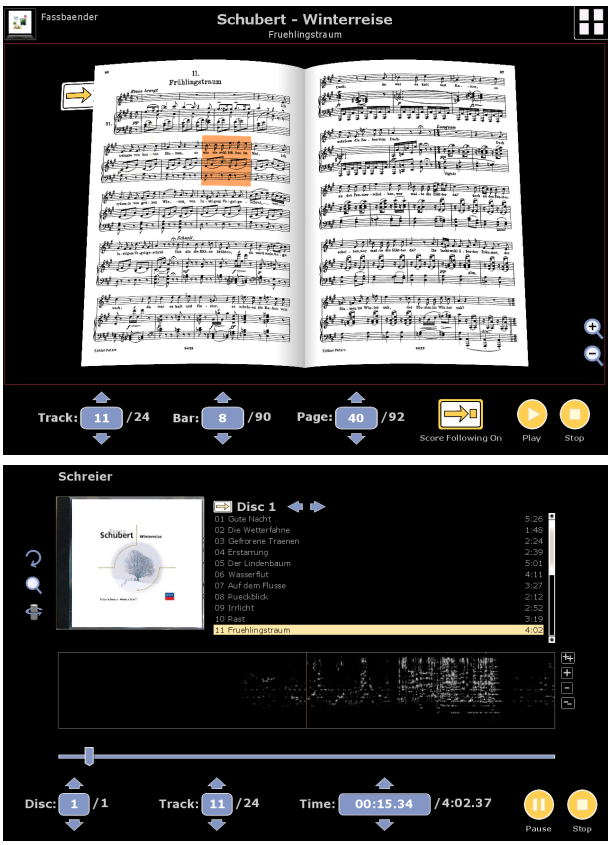

Figure 1: Screenshots of the user interface for displaying and interacting with sheet music books (top) and audio CD collections (bottom).

Cross-domain tasks

In the following, we assume that we are given a book of sheet music where each page is given by a scanned image and a collection of audio CDs where each track is given as an audio file. For each track, a first task is to find out which pages of the book correspond to which audio tracks. For example, audio track 3 might corrspond to pages 24 through 31. This leads to a mapping between sheet music pages and audio tracks. In practice, that mapping should be even finer grained, because track transitions may take place at any position within a single page of music notation. This mapping could be created manually, but especially for larger collections of music, a fully-automatic or semi-automatic approach is preferable.

Another cross-domain task is content-based retrieval in the case of sheet music and audio. Here, one task is finding sections in the audio database that match a given query of sheet music data or that are similar with respect to certain features of interest. Cross-domain retrieval would also cover the dual case of finding matches in a sheet music database given a query consisting of a block of audio data.

The goal of the task of cross-domain synchronization for sheet music and audio is to link regions within the two-dimensional image domain of sheet music documents to semantically corresponding temporal regions in audio recordings. One may choose one of several options for the granularity of the regions that are to be linked, for example, pages, lines, bars or notes. Depending on the choice of granularity, there are different strategies to perform the synchronization task. For example, in case of a page-wise synchronization the manual creation of links between the pages of sheet music and corresponding temporal regions of audio recordings may still be acceptable. However, aiming at finer grained resolutions, the effort

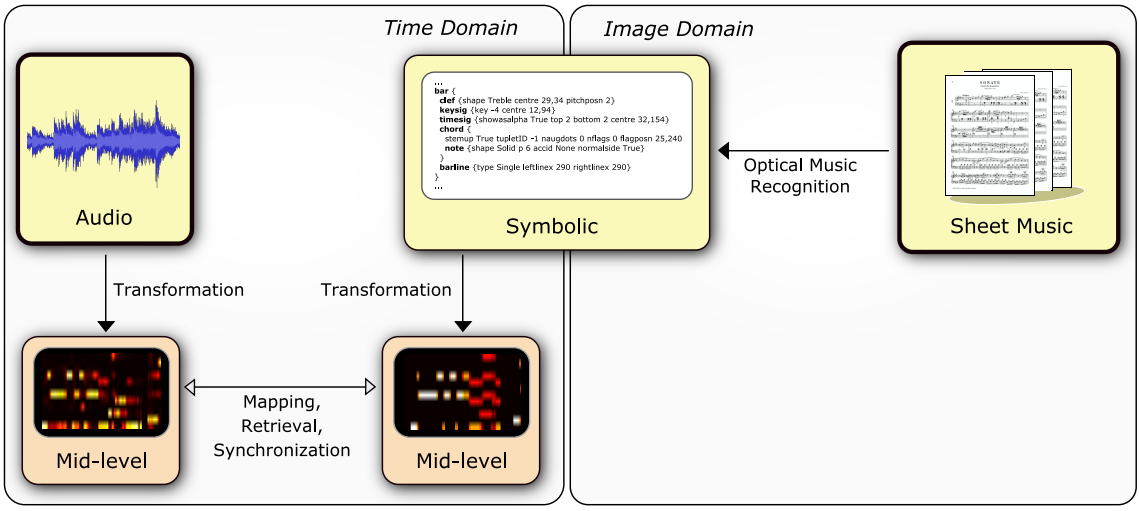

Figure 2: llustration of data types and data transformations relevant for our approach to performing cross-domain tasks with sheet music and audio.

required for creating the links increases significantly. Especially in the case of large document collections automated synchronization methods are indispensable.

Mid-Level Representation for Sheet Music and Audio

One key strategy of our approach to the implementation of cross-domain tasks as introduced in the previous section is to reduce the two different types of music data, the audio recordings as well as the scanned sheet music, to the same type of mid-level representation, which then allows for a *direct* comparison *across* the two domains, see Figure 2. In this context, chroma-based features have turned out to be a powerful mid-level music representation [1, 2, 3]. Here, the *chroma* correspond to the twelve traditional pitch classes of the equal-tempered scale and are commonly indicated by the twelve pitch spelling attributes $C, C^{\sharp}, D, \ldots$,B as used in Western music notation.

In the case of audio recordings, normalized chroma-based features indicate the short-time energy distribution among the twelve chroma and closely correlate to the harmonic progression of the underlying piece. Based on signal processing techniques, the transformation of an audio recording into a chroma representation (or chromagram) may be performed either by using short-time Fourier transforms in combination with binning strategies [1] or by employing suitable multirate filter banks [3].

The transformation of scanned sheet music into a corresponding chromagram requires several steps, see [4]. First, each scanned page is analyzed using optical music recognition (OMR) [5] to extract musical note parameters (onset times, pitches, durations) along with 2D position parameters from the scanned image. Using an estimate of the tempo, the explicit pitch and timing information can be used to derive a chromagram essentially by identifying pitches that belong to the same chroma class. A similar approach has been proposed in [2] for transforming MIDI data into a chroma representation.

Application and User-Interface

Figure 1 shows screenshots of the user interface developed for displaying and interacting with sheet music books and audio CD collections. Here, the sheet music and audio data are linked through cross-domain synchronization. During acoustic playback, the current playback position is being highlighted in the score book. The user can navigate to any position within the data by either using the track controls for the CD collection or by simply clicking on a bar in the score book. Cross-domain retrieval can be performed by selecting a range of bars in the score to query for similar regions in the audio data. The user interface and applications are utilized in the real-world scenario of the PROBADO digital library project [6].

References

[1] Bartsch, M.A., Wakefield, G.H.: Audio thumbnailing of popular music using chroma-based representations. IEEE Trans. on Multimedia **7**(1) (2005) 96–104.

[2] Hu, N., Dannenberg, R., Tzanetakis, G.: Polyphonic audio matching and alignment for music retrieval. Proc. IEEE WASPAA, New Paltz, US (2003).

[3] Müller, M.: Information Retrieval for Music and Motion. Springer (2007).

[4] Kurth, F., Müller, M., Fremerey, C., Chang, Y., Clausen, M.: Automated Synchronization of Scanned Sheet Music with Audio Recordings. Proc. ISMIR, Vienna, AT (2007) 261–266.

[5] Choudhury, G., DiLauro, T., Droettboom, M., Fujinaga, I., Harrington, B., MacMillan, K.: Optical music recognition system within a large-scale digitization project. Proc. ISMIR, Plymouth, MA, US. (2000).

[6] Diet, J., Kurth, F.: The Probado Music Repository at the Bavarian State Library. Proc. IMSIR, Vienna, AT (2007) 501–504.