GLOBE OF MUSIC - MUSIC LIBRARY VISUALIZATION USING GEOSOM

Stefan Leitich, Martin Topf
University of Vienna
Department of Distributed and Multimedia Systems
Liebiggasse 4/3-4, 1010 Vienna, AT

ABSTRACT

Music collections are commonly represented as plain textual lists of artist, title, album etc. for each contained music track. The large volume of personal music libraries makes them difficult to browse and access for users. In respect to possible information visualization techniques, no established convenient user interfaces exist. By using a spherical self-organizing map algorithm on low level audio features and processing the resulting map data, a Geographic Information System is used to visualize a music collection. This results in an aspiring music library visualization, which can be handled intuitively by the user and even provides new possibilities for accessing a music collection in the digital domain.

1 INTRODUCTION

The advent of digital media libraries has definitely passed. Nearly every media consumer is in possession of a digital library of media documents nowadays. This of course includes private music collections. In previous days, one browsed through her own music collection by picking CDs from shelves, rearranging stacks and heaps, while searching for the desired track. This was on the one hand an annoying task, but on the other hand somehow part of the listening experience. After having handled a CD a few hundred times, while searching for another one, the artwork as visual experience becomes strongly connected to the tracks the CD contains.

In the digital domain the dominant interface for browsing a personal media library was, and partly still is, a search task in a long list of plain textual information. Those lists contain known metadata of a song like artist, title, album, etc. In commercially available media playback applications, the textual playlist is still the most prominent kind of library representation that the user has to deal with. One exception is Apple iTunes with its “cover flow view”. In this visualization, the user has the possibility of browsing through the tracks of her media library, while the artwork is presented in a neatly animated horizontal lineup, with the selected song’s artwork in the center.

The coinciding popularity of Geographic Information Systems (GIS), for example Google Earth or NASA WorldWind [1], means that users are familiar with the interactions possible in such a system. One can easily navigate by rotating, tilting and zooming to access the desired view.

Our approach, *Globe of Music*, makes use of a spherical self-organizing map (SOM) algorithm, to arrange the items of a digital music library on a sphere in context of a GIS.

2 RELATED WORK

An early publication in the research domain of Music Information Retrieval (MIR) by Tzanetakis, tackled not only feature extraction aspects for audio and related classification methods, but also possible visualization techniques in this domain [9]. He proposed these techniques as interaction methods to enhance audio editors supported by automatic feature extraction, but also browsers for audio collections to visualize timbre or genre information of audio in a 2D and 3D manner.

Similar in structure to this approach is the work by Pampalk [7], who developed a psychoacoustically motivated approach for feature extraction and combined it with a visualization technique using a landscape metaphor - the so-called *Islands of Music*. This approach was further developed by Pampalk himself, using aligned SOMs to support multiple, user-adjustable views of a music collection [5], still using the landscape metaphor. Knees et al. [2] interpreted the height profile of the landscape metaphor to create a 3D representation of the *Islands of Music*. Additionally, they introduced an anisotropic auralization in this approach, by rendering a 5.1 surround sound model and augmenting the environment by related images and terms retrieved from the web in this user interface.

Contrasting visualization techniques are presented in the work by Torrens et al. [8]. Here, 2D visualizations in the form of a disc, a rectangle, and a tree-map are presented, using criterias like genre, artist, year of creation, and a definable quantitative criterion such as playcount.

Goto proposes in [1] an user interface - *Musicream* - combining the query-by-example and browsing paradigm.

[1] `http://worldwind.arc.nasa.gov`

Music tracks are visualized as music-discs and described by feature vectors. The so-called "streaming function" creates a steady flow of music-discs over the screen. If the user picks one of these discs and approaches another disc with it, similar music-discs, in terms of feature vector distance, will stick together and form a playlist. Another approach is the joined work by Goto and Pampalk [6] called *MusicRainbow*. This approach exploits a combination of signal based information (spectral similarity and fluctuation patterns) and information from the web. Visually, this depicts a "circular rainbow" with similar artists placed next to each other using a one-dimensional circular SOM. High-level terms (e.g."rock") on the inside and on the outside music-related terms (e.g. instruments, style attributes, etc.) are used for labeling. Colored rings encode different styles of music, and are sorted by the size of the styles part within the collection. A turning knob is used as an input device to browse the music collection.

Our approach is motivated by the landscape metaphor, but augments this with a spherical representation using a spherical SOM - *Globe of Music*. The spherical SOM intrinsically abolishes the border effect of a SOM, and the sphere or globe serves as the user interface metaphor of intuitive perception. A more detailed description is given in the next section.

3 APPROACH

Our approach is composed of three steps: (1) extracting signal-based features from audio, (2) using the GeoSOM algorithm [11] to arrange music tracks according to feature vectors, and finally (3) transforming this information into a GIS renderable format, which depicts the user interface.

3.1 Feature Extraction

The features used to describe the music pieces, are the Statistical Spectrum Descriptor (SSD) evaluated by Lidy in [4]. According to Lidy they showed reasonable performance for classification by similarity tasks in relation to feature vector dimension.

The SSD is composed of seven statistical moments (mean, median, variance, skewness, kurtosis, min- and max-value) for the 24 critical bands. The descriptor captures the frequency characteristics in terms of statistical information about the audio signal, resulting in a feature vector of 168 elements.

The collection used for initial user experiments (cf. Section 4) is the training part of the ISMIR Genre Classification Collection of the ISMIR 2004 contest, provided by Magnatune [2]. The collection used, consists of 728 tracks of different styles: classical (319), electronic (115), jazz/blues (26), metal/punk (45), rock/pop (101), world (122).

[2] `http://www.magnatune.com`

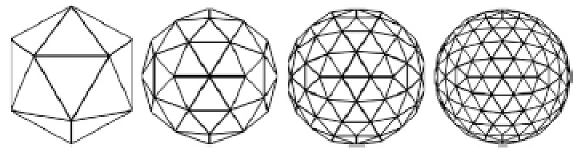

Figure 1. Icosahedron with tessellation frequency 1 to 4 [11].

3.2 GeoSOM

Neural networks, especially the self-organizing map proposed by Kohonen [3], are popular tools for information organization in visualization techniques used in MIR. In our approach we apply a spherical SOM, namely the GeoSOM proposed by Wu et al. in [11, 10], which provides, through a 2D data structure, a space- and time-efficient approach for neighborhood searching. For a detailed comparison of different spherical SOM implementations, refer to [11].

The GeoSOM is suited for the spherical representation through the use of a tesselated icosahedron (see Fig. 1) instead of a rectangular or hexagonal lattice of a basic SOM. Furthermore it reduces data distortion due to the removal of boundaries.

A short overview of this algorithm and its data structure is given in the following part of this paper. For a detailed description we refer to [11].

An icosahedron-based geodesic dome is, among the five platonic polyhedra, the most similar one to a sphere. With the exception of the original 12 vertices (5 neighbors), all vertices have 6 immediate neighbors. Depending on the desired number of neurons N ($N = f^2 * 10 + 2$), the icosahedron can be tessellated to increase its dome's frequency f.

Cutting the dome open and tilting in a way to retrieve an orthogonally aligned 2D grid, results in Fig. 2. The pole's (A, C) vertices are repeated 4 times (e.g. A, A', ... , A"") and vertices along the cut edges are duplicated once (e.g. E, E'). Be aware that there are also connections along one diagonal between two points (see enlargement of point E in Fig. 2). This must be considered when calculating the distance between two points in the 2D grid, and when searching and updating the neighborhood.

For storing the 2D matrix we used two one-dimensional arrays, called *u_array* and *v_array* as proposed in Wu et. al [11]. Vertices of a certain u coordinate are stored in a *v_array* (ordered according to their v coordinates) and those *v_arrays* themselves are contained in the u_array. The duplicated points can be found by starting from the marked points in Fig. 2 and proceeding in the indicated directions. Boundaries starting at points P and Q are duplicated as P' and Q' at the other end of the data structure.

The training process of the GeoSOM is similar to the training of a SOM, considering the different data structure regarding the tesselated icosahedron. After the training

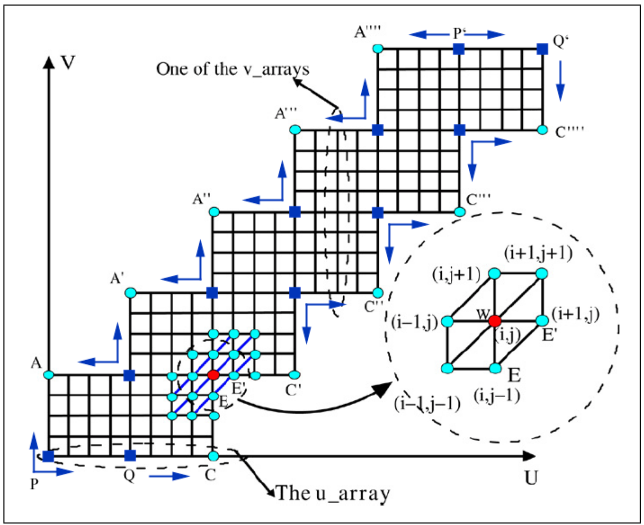

Figure 2. The 2D data structure for icosahedron-based geodesic domes [11].

Figure 3. Total view of *Globe of Music*.

process the GeoSOM data structure is mapped onto the sphere. The data structure is traversed in diagonals from the lower left to the upper right and coordinates are calculated.

3.3 Geographic Information System

We made use of the NASA World Wind (Ver. 1.4) GIS, an open source project developed at NASA Ames Research Center. It allows users to explore the earth via satellite imagery (e.g. LandSat dataset), elevation data, and meteorologic data, also to retrieve views of other planets. World Wind offers the possibility of extending its functionality via a plugin mechanism and the definition of custom "worlds". For an extensive description of NASA World Wind we recommend the project's website [3].

In order to define a custom globe like the *Globe of Music*, an XML file defining the most basic attributes of a world - for example name, equatorial radius and server connection information, if the world contains remotely stored terrain data - has to be created

Fig. 3 shows a screenshot of the *Globe of Music* visualization in World Wind.

Information displayed on a globe in World Wind is arranged in layers above the basic terrain data. Layers make it possible to place, icons on certain longitude/latitude coordinates with a defined height above surface, including name, texture, dimension, and a link, to be followed when clicking the related icon. These are the most important attributes used in our implementation, the possibilities for visualizing information on a layer are manifold.

In a layer definition the data defined in boundaries and placename files can also be imported by referencing the boundary and placename data, and defining visualization parameters such as the range of viewers altitude for visibility, distance above surface, font, etc.

The layer we have created contains a background image and icons for every music track placed on the *Globe of Music*. Icons are textured with the artwork and linked to the corresponding music file (see Fig. 3). Pointing to an icon with the mouse cursor displays the track's artist and title information.

Coordinates of the icons are determined by the neuron (sector) of the GeoSOM the music track belongs to. If there is more than one track placed on a neuron, tracks are stringed in this sector next to each other on the globe.

Boundaries (e.g. a country's border) and placenames (e.g. a country's name) are represented in World Wind via binary file formats. With the *Globe of Music*, these formats are used to (1) outline sectors on the globe related to a GeoSOM neuron, and (2) to define placenames, in our case artist name and track title, for the coordinates at which icons are placed. This allows a visualization of the sectors that the globe is divided into, and the usage of the built-in place finder tool to query for and locate individual tracks by text search.

4 INITIAL USER EXPERIMENTS

In the setting of the initial user experiments, users were presented the *Globe of Music* visualization and were given a description and mission in the form of: (1) This visualization depicts a music collection. (2) Music tracks are arranged on the globe, represented by icons of related artwork. (3) Minor interaction explanations were given if required (rotate, zoom, tilt, click for playback). (4) The music collection contains music tracks presumably unknown to the user. (5) Browse through and explore this collection and try to determine its composition.

Test-users were allowed to browse for approximately 15-20 min., followed by a survey about their experience. The composition of the group of test-users was in terms of gender 2 to 1 (male to female). The group's size was 12 people, with an average age of 27.25 ($\sigma \approx 5.88$).

A survey was conducted on the basis of a questionnaire and a semi-structured interview, by asking questions about the composition of the music collection involved, organi-

[3] `http://worldwind.arc.nasa.gov`

zation of the music title's placement, experiences of orientation in the visualization, and letting the test-users comment freely.

The test candidates were asked to guess the size of the music collection presented, and to enumerate the varieties of music genres, as well as the size of a genre's collection part found while browsing. The intention was to figure out how well an unknown music collection can be explored. The genres contained in the collection could be clearly determined, 10 out of 12 persons found all of them. Classical music was correctly determined as the dominant genre of the collection.

Users were asked about their impression of the organization of songs on the globe. They were not given any information a priori, neither did the visualization include any information about organization of the placement. The only way to determine the organization was by listening to the music tracks. Again, 10 out of 12 persons were convinced there is a systematical approach in the placement of tracks. Common agreement was present about the neighborhood placement of similarly perceived songs, but speculations varied from longitudinal or latitudinal mapped parameters for placement.

The determination of orientation in the *Globe of Music* was done by asking the test-users to locate a music title they recalled whilst browsing. This task was performed very well and proved the importance of visual association between music listened to and its artwork. The greater part of the test group stated it was easy to find a certain track by remembering the related artwork.

In common comments, users reported a problem with navigation, because of the globe's axis. This circumvents the free rotation around all axes and makes navigation difficult near the poles. Overall, users were very interested in using such a visualization and asked for common playback application features such as managing playlists, adding new songs to the collection, and information about the actual music title playing.

5 CONCLUSION

In this paper we presented a novel approach for visualizing music collections using a GIS metaphor. It was conducted by feeding a spherical SOM (GeoSOM) with signal based feature vectors for music track description, and visualizing the resulting spatial distribution on a globe in NASA World Wind. Inital user experiments with our prototype showed promising results in terms of high user acceptance. This visualization provides the user with an intuitive interface which is easy to handle and fun to explore. The ideas and feature requests collected during experiments encourage further development.

6 ACKNOLEDGEMENTS

We would like to thank the students Robert Neuner, Andreas Seidler, and Michael Müller supporting our work with their investigations regarding data formats and extending NASA World Wind.

7 REFERENCES

[1] M. Goto and T. Goto. Musicream: New music playback interface for streaming, sticking, sorting, and recalling musical pieces. In *Proceedings of the 6th International Conference on Music Information Retrieval*, pages 404–411, London, UK, 11-15 September 2005.

[2] P. Knees, M. Schedl, T. Pohle, and G. Widmer. An innovative three-dimensional user interface for exploring music collections enriched. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 17–24, New York, USA, 2006. ACM Press.

[3] T. Kohonen, editor. *Self-organizing maps*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1997.

[4] T. Lidy and A. Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In *Proceedings of the 6th International Conference on Music Information Retrieval*, pages 34–41, London, UK, 11-15 September 2005.

[5] E. Pampalk, S. Dixon, and G. Widmer. Exploring music collections by browsing different views. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR'03), Washington, D.C., USA*, October 2003.

[6] E. Pampalk and M. Goto. Musicrainbow: A new user interface to discover artists using audio-based similarity and web-based labeling. In *7th International Conference on Music Information Retrieval*, pages 367–370, Victoria, BC, Canada, 8-12 October 2006.

[7] E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 570–579, New York, USA, 2002. ACM Press.

[8] M. Torrens, P. Hertzog, and J.L. Arcos. Visualizing and exploring personal music libraries. In *5th International Conference on Music Information Retrieval*, Barcelona, Spain, 10-14 October 2004.

[9] G. Tzanetakis. Manipulation, analysis and retrieval systems for audio signals. Technical Report TR-651-02, Princeton Computer Science, 2002.

[10] Y. Wu and M. Takatsuka. The geodesic self-organizing map and its error analysis. In *ACSC '05: Proceedings of the Twenty-eighth Australasian conference on Computer Science*, pages 343–351, Darlinghurst, Australia, 2005. Australian Computer Society, Inc.

[11] Y. Wu and. Takatsuka. Spherical self-organizing map using efficient indexed geodesic data structure. *Neural Networks*, 19(6-7):900–910, July-August 2006.