Contents Analysis and Synthesis Scheme for Music Album Cover Art

Daejin Moon*, Seungmin Rho*, Eenjun Hwang*★

Abstract

Most recent web search engines perform effective keyword-based multimedia contents retrieval by investigating keywords associated with multimedia contents on the Web and comparing them with query keywords. On the other hand, most music and compilation albums provide professional artwork as cover art that will be displayed when the music is played. If the cover art is not available, then the music player just displays some dummy or random images, but this has been a source of dissatisfaction. In this paper, in order to automatically create cover art that is matched with music contents, we propose a music album cover art creation scheme based on music contents analysis and result synthesis. We first (i) analyze music contents and their lyrics and extract representative keywords, (ii) expand the keywords using WordNet and generate various queries, (iii) retrieve related images from the Web using those queries, and finally (iv) synthesize them according to the user preference for album cover art. To show the effectiveness of our scheme, we developed a prototype system and reported some results.

Key words: album cover art, music mood, keyword extraction, wordnet, content-based retrieval

I. Introduction

Due to the popularity of various media compression formats such as mp3, traditional music players are being rapidly replaced by versatile media players with screens. The screen is usually used for displaying various information including singer, music title, and album cover art, etc. Basically, album cover art contains several core pieces of information about the music including images associated with the mood, place, time, singer and album title. As a result, album cover art is a very effective means for visualizing and advertising music. Conventional music albums or music compilation album cover art are usually made by music contents providers, music contents manufacturers and professional cover art designers. Hence, the creation of such album cover art is very difficult for non-professional users who personally make their own music compilation albums using podcasts.

For automatic generation of such music album cover art, many issues should be addressed. First, since the cover art needs to suggest the music mood, this should be extracted from the music automatically based on low-level signals and acoustic features [11-13] or musical features [14, 15]. Also in [15] the authors considered musical mood variation for automatic mood detection and tracking.

Secondly, music lyrics can contain very useful information about music such as mood, place and time, and hence should be investigated. Meaningful words are extracted from the music lyrics and their implicit moods should be inferred. There have been many works on this in the text mining area[5, 23].

Thirdly, due to the availability of various SNS (Social Network Service) sites, various multimedia contents are now available with voluntary annotations including place, time and emotional features. Especially, recent works on music contents [3] have revealed that user collaboration and

*School of Electrical Engineering, Korea University

★ Corresponding author

※This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0025395)

Manuscript received Dec. 3, 2010 ; revised Dec. 29, 2010

SNS-based web services can provide better music retrieval performance than traditional content-based analysis.

In this paper, we present a new method for automatically generating an album cover art for a list of music. The album cover art is an image representing the overall music mood, background place, time, and some other metadata including singer and music title. To achieve this, we carry out three different music analyses and combine the result. First, we perform SMERS-based music mood extraction [6]. Second, we extract the principal keywords from the music lyrics and refine them through WordNet in order to determine their major mood and context information. Third, we analyze music metadata such as ID3 tag information in the MP3 file. These keywords are used as query input to retrieve relevant images on the web. Those retrieved images are synthesized with metadata according to the user preference, by using various artistic effects to generate the album cover art.

The outline of this paper is follows. In Section 2, we describe music mood extraction based on acoustic/musical features and keyword analysis of music lyrics. In Section 3, we present a means to generate music album cover art automatically based on music mood and relevant metadata. In Section 4, we briefly describe our prototype system and present some experimental results. In Section 5, we conclude this paper.

II. Related works

1. Music mood extraction

To date, many works have attempted to establish models of emotions and factors leading to the perception of such emotions in music. Traditional mood and emotion research in music has focused on finding psychological and physiological factors that influence emotion recognition and classification. During the 1980s, several emotion models were proposed, which were largely based on the dimensional approach to emotion rating.

The dimensional approach focuses on identifying emotions based on their location in a small number of dimensions such as valence and activity. Russell's[18] circumflex model has had a significant effect on emotion research. This model defines a two-dimensional, circular structure involving the dimensions of activation and valence. Within this structure, emotions that are on opposite sides of the circle, such as sadness and happiness, correlate inversely. Thayer [16] suggested a two-dimensional emotion model that is a simple but powerful means of organizing different emotion responses: stress and energy. The dimension of stress is called valence while the dimension of energy is called arousal.

As shown in Figure 1, the two-dimensional emotion plane can be divided into four quadrants with eleven emotion adjectives superimposed. We use eleven types based on Juslin's theory and Thayer's emotion model.

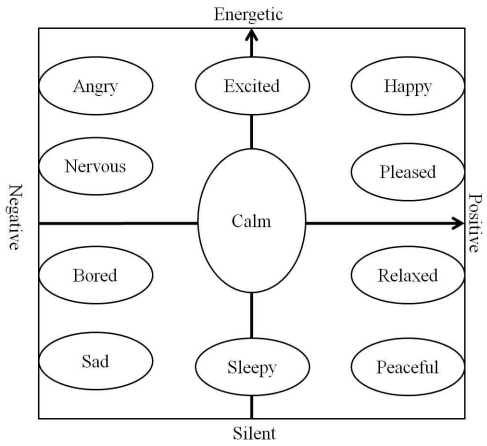

Fig. 1. Thayer's 2-dimensional emotion model

Other works have attempted to investigate the influence of music factors such as loudness and tonality on the perceived emotional expression [16, 17]. The authors analyzed those factors using diverse techniques, some of which involved measuring psychological and physiological correlations between states of particular musical factors and emotion evocation. According to [17], Juslin and Sloboda investigated the utilization of acoustic cues in the communication of music emotions between performers and listeners, and measured the correlation between emotional expressions (e.g., anger, sadness and happiness) and acoustic cues (e.g., tempo, spectrum and articulation).

Automatic emotion detection and recognition in speech and music is growing rapidly due to technological advances in digital signal processing and various effective feature extraction methods. Emotion recognition can play an important role in many other potential applications such as music entertainment and human-computer interaction systems.

One of the first studies on emotion detection in music was presented by Feng et al. [15]. Their work, based on Computational Media Aesthetics (CMA), analyzes the two dimensions of tempo and articulation, which are mapped onto four mood categories: happiness, anger, sadness and fear. Lie et al. [14] developed a hierarchical framework for extracting music emotion automatically from acoustic music data. They used music intensity to represent the energy dimension of the Thayer model, and timbre and rhythm for the stress dimension.

2. Lyric analysis

Music lyrics consist of a finite number of words, some of which are chosen to represent the meaning of the music or music mood. Thus, identifying such words can be useful for music mood extraction and classification. In [4], the authors calculated various properties including (i) tf-idf (*Term Frequency - Inverted Document Frequency*)weight based on the frequency of lyric words, (ii) existence of such words so as to represent specific moods, and (iii) frequency of words in a specific part-of-speech. The first property showed the best accuracy.

WordNet [8] is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The purpose is twofold. First, it provides a combination of dictionary and thesaurus for more intuitive utilization, and second, it supports automatic text analysis and artificial intelligence applications. The database and software tools have been released under a BSD style license and can be downloaded and used freely. The database can also be browsed online.

3. Metadata

To date, metadata has been useful in the text-based retrieval of multimedia contents. For instance, title and singer are representative metadata for text-based music search. In this paper, we use the ID3 tag information of the mp3 file as its metadata. An ID3 tag is a metadata format for an mp3 file. Various information on the music such as title, singer, and album name is included.

III. Automatic Music Album Cover Creation

1. Overall Structure

Figure 2 shows the overall structure for creating music album cover art automatically. First, for the music list to create album cover art, we determine relevant keywords based on (i) SMERS-based music mood extraction [6], (ii) music lyrics clustering based on a semi-supervised clustering method, and (iii) ID3 tag analysis contained in the music content file. Typical keywords include music mood, place, time, and singer information. Those keywords are refined by query expansion and query reformulation in order to retrieve relevant images from Google images and Flickr. Finally, retrieved images are synthesized with relevant metadata according to the user preference to generate the music album cover art.

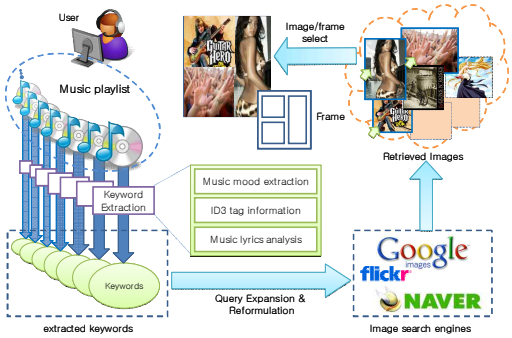

Fig. 2. Overall system structure

2. Music Mood Extraction

For automatic music mood extraction, we used the SMERS [6] system. This system consists of three main modules: (i) The feature extraction module extracts and analyzes seven distinct musical features, viz., pitch, tempo, loudness, tonality, key, rhythm and harmonics; (ii) The feature-emotion mapping module maps those extracted features onto

eleven emotion categories based on Thayer's two-dimensional emotion model (See Figure 1); (iii) The training module trains the SVR (Support Vector Regression) using the extracted features as input vectors. We used two distinct SVR functions for predicting arousal and valence values based on acoustic features in a polar coordinate system. One is for the distance from the origin to the emotion in a Thayer-like coordinate system, and the other is for the angle. Using these two trained SVRs, the tool recognizes each song's emotion. An empirical test shows that the polar coordinate system gives better results than the Cartesian coordinate system [6].

3. Keyword Extraction form Lyrics

In order to extract keywords from music lyrics, we use public lyrics data available at Lyricwiki.org. The lyrics gathered on the web contain multiple types of noise. To remove such noises, we performed the preprocessing steps described in [5]. Most lyrics have sections such as *intro, interlude, verse, chorus* and *outro*. These annotations should be removed because they are not important. However, repetition words such as *[repeat 2], ×5* are important. Hence, we remove such repetition words from the lyrics and repeat the applicable lyrics as many times as needed.

After removing noises, we perform the stemming process on the lyrics in order to reduce inflected words to their stem, base or root form. We used the Porter stemming algorithm [21]. Prior to the next step, we also remove stopwords such as *a, an, as, in, the, to, etc*[7].

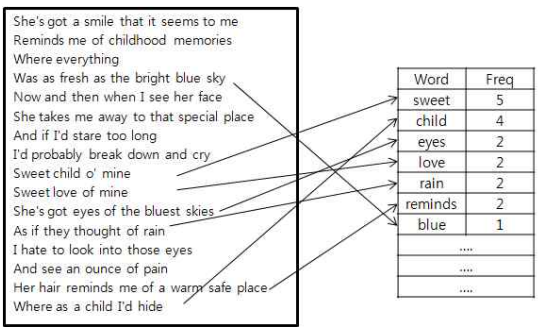

Fig. 3. Example of bag-of-words model

After the stemming process, the lyrics are segmented into words based on the bag-of-words model [5]. As shown in Figure 3, the bag-of-words model is commonly used in natural language processing and information retrieval. It represents a text such as a sentence or document by an unordered collection of words, regardless of relationships such as grammar and word ordering.

The keywords returned by the bag-of-words model are those which appear most frequently in the lyrics. Not all these keywords are meaningful for the album cover art. In order to remove such meaningless keywords, we use query expansion and reformulation techniques for filtering English adjectives and nouns. For instance, in the All Music Guide (AMG) [11] that was classified by music professionals and widely used as metadata in the Music Information Retrieval (MIR) field, it was observed that some music could have as many as 180 different moods simultaneously. In this paper, we filter such keywords using WordNet [8], which is a lexical database for English words. We classify keywords into several subjects using the lexical file information of WordNet. We use keywords belonging to subjects that are relevant to the context information such as time, location, and feeling for image retrieval.

4. Metadata Analysis

Usually, a normally circulated sound file contains metadata such as an ID3 tag [19]. This tag contains many types of detailed information on the music including song title, artist, album name, composer, conductors, media types, BPM, lyrics, etc. There are two unrelated versions of ID3: ID3v1 and ID3v2, which are incompatible and can exist in a single file simultaneously. However, not all MP3 files have the same amount of metadata. So we utilize keywords for artist, song title, and album name through an ID3 tag, since these are usually available. We transform this tag information into the bag-of-words model. These words are not dependent on the frequency, and can be classified into semantic keywords again through WordNet. Finally, the classified keywords are used in image retrieval together with the keywords from the song lyrics.

5. Image Retrieval

Based on the extracted keywords, we collect relevant images from the web using traditional web

image search engines such as Google and Flickr. For image retrieval, Google and Flickr provide open API methods. Depending on the properties of the keywords used, we select appropriate images as candidate images for the album cover art.

IV. Experimental Results

1. Experimental setup

To show the effectiveness of our scheme, we implemented a prototype system on Windows Vista Enterprise. The web server was Apache 2.2 with PHP 5 and MySQL5.

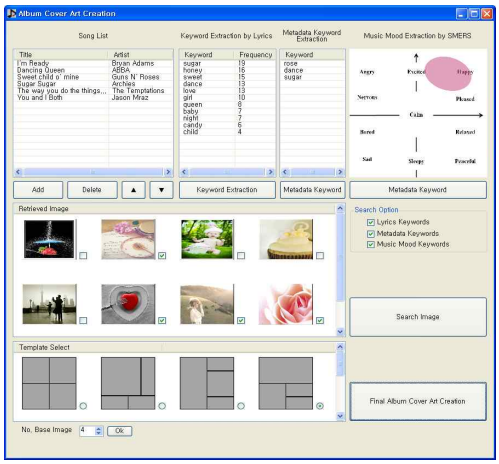

Fig. 4. Automatic album cover art creation test program

Fig. 5. Sample final album cover art

Also, image retrieval was based on the Flickr [9] and Google [10] API. For the experiment, we used music files in the MP3, WMA, or AAC format with tags that contained some metadata.

2. Album cover art creation

Figure 4 shows the user interface of our prototype system for generating album cover arts for a sample music list. The user interface consists of six major parts: i) song list, ii) keyword list for the extracted lyrics, iii) keyword list extracted from ID3 tag information, iv) music mood extractor, (v) cover art template selector and (vi) image selector. For a music list selected from a music database, their keywords are extracted by music mood extraction, lyrics analysis, and ID3 tag information analysis.

Table 1. Example of song list

Title	Artist
I'm Ready	Bryan Adams
Dancing Queen	ABBA
Sweet Child O' Mine	Guns N' Roses
Sugar Sugar	Archies
The Way You Do The Things You Do	The Temptations
You and I Both	Jason Mraz

For instance, for the sample music list shown in Table I, Table II shows the keywords including lyrics, metadata and music mood that are used for web image search. A sample album cover art generated according to the selected template is shown in Figure 5.

Table 2. Example of keywords list

Lyrics Keywords		Metadata Keywords	Music Mood
Term	Freq		
sugar	19	rose	happy
honey	16	dance	
sweet	15	sugar	
dance	13		
love	13		
girl	10		
queen	8		
baby	7		
night	7		
candy	6		
child	4		

At present, the system provides quite simple templates. However, for a more satisfactory result, we plan to add much more complicated and artistic templates.

V. Conclusion

In this paper, we proposed a novel scheme for automatically generating music compilation cover art for a music list. In order to collect candidate images for cover art from the web, we collected keywords through three different music analysis methods: (i) music mood extraction from music based on acoustic features, (ii) meaningful keyword extraction from music lyrics, (iii) ID3 tag information analysis including singer, title and album name. We implemented a prototype system and showed that it can definitely generate album cover art effectively. At present, the synthesis templates of candidate images for album cover art are simple. To make it more professional, we are making more artistic templates.

References

[1] Google, http://www.google.com/
[2] Flickr, http://www.flickr.com/
[3] J. H. Kim, B. Tomasik, D. Turnbull, "Using artist similarity to propagate semantic information," *10th International society for music information retrieval conference* (*ISMIR 2009*), pp.375-380, Oct. 2009
[4] R. Mayer, R. Neumayer, and A. Rauber, "Rhyme and style features for musical genre categorisation by song lyrics," *ISMIR 2008* pp.337-342, 2008
[5] X. Hu, et al., "Lyrics text mining in music mood classification," *ISMIR 2009*, pp.411-416, Oct. 2009
[6] Byeong-jun Han, Seungmin Rho, Roger B. Dannenberg, Eenjun Hwang, "SMERS: Music emotion recognition using support vector regression," *ISMIR 2009*, pp.651-656, Oct. 2009
[7] stopwords, http://en.wikipedia.org/wiki/Stop_words/
[8] WordNet, http://wordnet.princeton.edu/
[9] Flickr API, http://www.flickr.com/services/api/
[10] Google API, http://code.google.com
[11] A. Ghias, et al., "Query by humming - musical information retrieval in an audio database", *in Proceedings of ACM Multimedia 95*, 1995,pp.231-236
[12] Hoashi, Zeitler, Inoue, "Implementation of relevance feedback for content-based music retrieval based on user preferences", *ACM SIGIR 2002*, pp.385-286, 2002
[13] Hoashi, Matsumoto, Inoue, "Personalization of user profiles for content-based music retrieval based on relevance feedback", *ACM Multimedia 2003*, pp.110-119, 2003
[14] Lie Lu, Dan Liu, Hong-Jiang Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals", *IEEE Transactions on Audio, Speech and Audio Processing*, vol.14, no.1, Jan.2006, pp. 5-18
[15] Yazhong F,Yueting Z, Yunhe P, "Music information retrieval by detecting mood via computational media aesthetics", *IEEE/WIC International Conference on*, pp235-241, Oct. 2003
[16] R. E. Thayer: "The Biopsychology of Mood and Arousal," *NewYork: Oxford University Press*, 1989
[17] P.N. Juslin and J.A. Sloboda: "Music and Emotion: Theory and research," *Oxford Univ. Press*, 2001
[18] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol. 39, 1980
[19] ID3, http://en.wikipedia.org/wiki/ID3/
[20] Smola, Alex J., et al., "A tutorial on support vector regression," *Statistics and computing*, vol. 14, pp.199-222, 2004
[21] M. F. Porter, "An algorithm for suffix stripping", *Program*, 14(3):130 - 137, 1980
[22] D. Yang, and W. Lee: "Disambiguating music Emotion Using Software Agents," *In Proceedings of the 5th International Conference on Music Information Retrieval* (*ISMIR'04*), 2004.
[23] Yajie Hu, Xiaoou Chen and Deshun Yang, "Lyric-based song emotion detection with affective lexicon and fuzzy clustering method," *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009

BIOGRAPHY

Daejin Moon (Student Member)

He received his B.S. degree in Computer Information Engineering from Dongseo University, Korea, in 2009. Currently he is pursuing the M.S. degree in the School of Electrical Engineering in Korea University. His research interests include text mining, multimedia search and telematics. Text mining, multimedia search and telematics.

Email : wizardyk@korea.ac.kr

Seungmin Rho (Member)

Seungmin Rho received his MS and PhD Degrees in Computer Science from Ajou University, Korea, in 2003 and 2008, respectively. In 2008-2009, he was a Postdoctoral Research Fellow at the Computer Music Lab of the School of Computer Science in Carnegie Mellon University. He is currently working as a Research Professor at School of Electrical Engineering in Korea University.

His research interests include database, music retrieval, multimedia systems, machine learning, knowledge management and intelligent agent technologies.

Email : smrho@korea.ac.kr

Eenjun Hwang (Member)

Eenjun Hwang received his BS and MS Degree in Computer Engineering from Seoul National University, Seoul, Korea, in 1988 and 1990, respectively; and his PhD Degree in Computer Science from the University of Maryland, College Park, in 1998. From September 1999 to August 2004, he was with the Graduate School of Information and Communication, Ajou University, Suwon, Korea. Currently he is a member of the faculty in the School of Electrical Engineering, Korea University, Seoul, Korea. and Communication, Ajou University, Suwon, Korea. Currently he is a member of the faculty in the School of Electrical Engineering, Korea University, Seoul, Korea.

His current research interests include database, multimedia systems, audio/visual feature extraction and indexing, semantic multimedia, information retrieval and Web applications.

Email : ehwang04@korea.ac.kr