Cross-Media Retrieval for Music by Analyzing Changes of Mood with Delta Function for Detecting Impressive Behaviours

Yoshiyuki Kato
Faculty of Environment and Information Studies
Keio University
5322 Endo, Fujisawa, Kanagawa, Japan
t10247yk@sfc.keio.ac.jp

Shuichi Kurabayashi
Faculty of Environment and Information Studies
Keio University
5322 Endo, Fujisawa, Kanagawa, Japan
kurabaya@sfc.keio.ac.jp

***Abstract*—This paper proposes a system that retrieves music by accepting a sequence of images as a query representing a change of sentiments in music. The system offers a query model that utilizes image files as a media for describing users' emotional demands for continuous changes of mood in music. This query model offers two types of delta functions, corresponding to music and images. Each delta function measures continuous changes in the corresponding sequential media. Applying the delta functions to the media data generates the values representing changes of moods. Each delta value is normalized distance in the corresponding metric space, thus, comparison of delta values extracted from heterogeneous media data makes it possible to calculate the cross-media relevance score. As a prototype implementation, we have developed a Web-based cross-media retrieval engine that provides an integrated user interface (UI) to create music queries by novices who may submit only rough and simple information.**

Keywords-Cross Media Retrieval; Multimedia Database; User Interface; Mood Analysis;

I. INTRODUCTION

Although music is a very common media in our daily lives, it is difficult to find music satisfying our preferences. This is because music changes in sentiments with time. In order to find the desired type of music, a user must listen to several parts of music in repositories, such as online music stores and personal music players. The main gateways for those online stores are provided as Web-based services, such as Amazon MP3 and Google's Play Music service.

In spite of the fact that young age users tend to select music according to their feelings, users have serious frustration to retrieve their favorite music, due to the lack of web technologies for inputting queries to find continuous media data such as music. Especially for finding recently released music that is unknown to a user, a method to describe a user's ambiguous desires for music is required [1]. Novices need a toolkit that assists its users to form their own queries in a trial-and-error manner. It is desirable to develop an intuitive Web-based toolkit for representing the demands of users for music.

Toward the above objective, this paper proposes a web-based music retrieval method that offers a cross-media query model utilizing "image files" as a media for describing users' emotional demands for continuous changes of mood in music. It is important to develop a stream-oriented query construction method for music, because music changes its content and impression with time. Our query model interprets the perceptional effects of temporal changes in media features such as tonality in music and colors in images. This paper shows a prototype system implementation realizing web-based music retrieval with considering changes of mood.

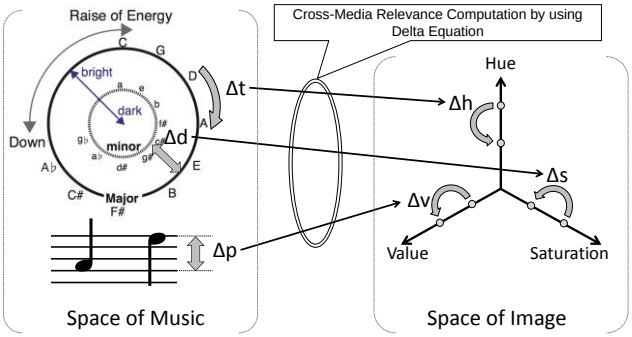

Figure 1. An Conceptual Overview of the Cross-Media Retrieval by using Delta Equation on Each Media Data

Our method achieves the cross-media retrieval by comparing the results of continuous sentiment analyses of music and images. In order to analyze the temporal changes in media data, this system provides two types of delta functions, corresponding to music and images, to generate the sequential values representing changes of mood. Those sequential values represent how the media data changes its sentiment with time. The system calculates the sentiment-oriented relevance score of music and images by comparing the calculated delta values.

Our design principle for this system is to make it possible to search musical contents that are invisible to users by using visible image contents. This concept is highly suitable for the web-based music retrieval because web is a visual media. Thus, a key technology of this system is a cross-media query interpretation method that recognizes how the media changes with time by using several metric spaces to calculate distances between the states of the media and the previous states of the media. For music, we have implemented three metrics based on music tonality analysis methods [2][3]: a tonality metric, a pitch metric, and a major/minor metric. Corresponding to these metrics for music, we have implemented three

 ISBN: 978-1-61208-280-6

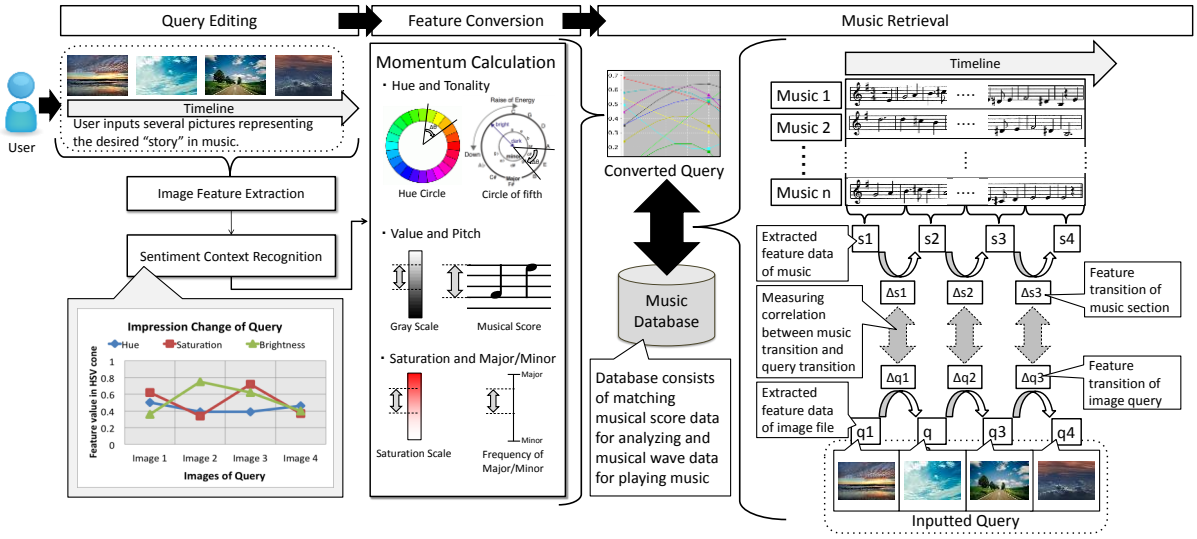

Figure 2. System Architecture for Retrieving Music by Delta Function Analyzing Changes of Mood

metrics for images based on the HSV color space: a hue metric, a saturation metric, and a value metric.

The system transforms invisible changes in the impression of music into visible changes in color and vice versa. Our approach converts "a delta value", which is distance in each space, between two spaces rather than a feature value itself because the system focuses on how the changes of mood affect on human perceptions. The most important feature of the two metric spaces is their configuration as topologically equivalent structures (Figure 1). Each axis in music space has a corresponding axis in the color space. Specifically, tonality is associated with hue, pitch with value, and major/minor with saturation. Thus, a specific distance in music space can be converted into the same distance in color space.

The advantage of this system is an intuitive method for users to edit a query in a trial-and-error manner, depending on their impression. The method makes it possible for users to describe changes in impression of music, which is difficult to represent directly, as a sequence of images with visually enhanced user interface, wherein the order of the images represents a change of impression.

The remainder of this paper is structured as follows. Section II presents motivating example of our query processing. Section III briefly summarizes the related work. Section IV describes the fundamental concept and the system architecture. Section V concludes this paper.

II. MOTIVATING EXAMPLE

In this section, we present a motivating example of our stream-oriented cross-media retrieval. In this example, a user wants to find a music that initially gives a dark impression and gradually makes a brighter impression like a sunrise, but the user does not have clear idea about the title and the artist of the music. In this case, by using our system, the user can retrieve the music with combining four pictures as shown in the left-hand side of Figure 2. The horizontal axis from left to right corresponds to elapsing time. This query represents changes in impression as follows: brightness (value in HSV color model)" is gradually increasing and then decreasing, and hue (type of colors) is stable, and saturation (vividness of colors) changes drastically with time. Hence, according to the relation between a pair of the dual-metrics, the system retrieves music corresponding to changes in impression as follows: pitch is gradually increasing and then decreasing, tonality is stable, and major/minor changes drastically.

III. RELATED WORK

The music information retrieval (MIR) system is a well-known means of helping users to find music by using several intuitive queries [1][4]. A traditional example of the MIR method is a content-based MIR system [5]. As a typical method of the content-based music information retrieval system, query by humming (QBH) makes it possible to use humming as a search key for finding a musical composition [6]. QBH is effective to find music that is familiar to a user, but cannot be applied to find music that is unknown to the user. As described in [7], novel techniques to search for new music that we have not heard are in great demand, because the conventional content-based music retrieval methods do not satisfy such queries.

Ciuha et al. [8] show a music visualization system that utilizes cross-media relationship of colors and tonalities. This system partially supports users in finding new and unknown music. Multi-timescale visualization techniques for displaying the output from key-finding algorithms were presented in [9]. An impression-based music visualization method that utilizes a result of a synesthesia study [10] was proposed in [11]. This uses a color sense of tonality to view the harmonic structure and relationships between key regions in a musical composition.

The most significant difference between the conventional approaches and our approach is that our system focuses on metrics implied by feature transitions in vector space, as the elements representing changes in the impressions of media. We do not use a knowledge base of music and color referring to synesthesia since this

 ISBN: 978-1-61208-280-6

is difficult to personalize for each user by reflecting their individual differences in impressions of media.

IV. System Architecture

In this section, we present our music retrieval system that interprets an image sequence as changes in the sentiment of music. As shown in Figure 2, the system consists of three main components as follows: 1) a query editor, 2) a feature conversion module, and 3) a retrieval engine. The query editor is the front-end module of the system. This module provides a set of operations to prepare and modify image files as a query according to a user's preferences. For example, the system implements an image-editing operator equipped with several color filters to change the overall impression of the image. The core component of the system is a retrieval engine that calculates the correlation between the query and retrieval target items according to their own changes in impression by using the metadata generated in the feature conversion module.

The system models the concept of "change in sentiment" by measuring the distance caused by feature transition of media data. Specifically, the system provides a bridging mechanism between the musical tonality metric space and the HSV color metric space. The bridge converts a distance calculated in the music space into a distance in the image space as keeping its impression factor. For example, in the distance conversion mechanism, hue, which is a type of color, corresponds to tonality, which is a type of music structure. By converting distances between heterogeneous metric spaces, the system realizes cross-media retrieval for stream media such as music.

A. Data Structure

The data structure in this system consists of two data elements, which are the image query and the music. An image-query object Q is defined as follows: $Q := \langle\langle h_1, s_1, v_1\rangle \ldots \langle h_i, s_i, v_i\rangle\rangle$, where h_i is hue data, s_i is saturation data, and v_i is brightness data, in the *i*-th image of the query. The system converts the RGB color values of each image into HSV triples at the pixel level. We define the metric space of images as the HSV color metric space with three axes: hue, saturation, and value. These three elements are significant factors affecting the impression of the image.

Hue represents the differences of color phases such as red, yellow, green, and blue. In the HSV cone of images, hue is represented by angle. The system converts the extracted hue angle in HSV cone of images into a hue scalar *h*, which is a value between 0 and 1. Saturation is vividness of color. In our system, the saturation is an average value of the vividness in an image. The system processes this vividness value of an image into the saturation scalar *s*, which is a value between 0 and 1. Here 0 and 1 represent the lowest value and the highest value, respectively. Value is brightness of color. The prototype system calculates the value (brightness) as an average value of color brightness in an image. Our system processes this brightness value of an image into a value of brightness scalar *v*, which is a value between 0 and 1.

When the proposed system receives a query consisting of several images, the proposed system divides music into sections, and the number of sections is equal to the number of images inputted as the image query. Thus, in this paper, a music object M is defined as follows: $M := \langle\langle t_1, d_1, p_1\rangle \ldots \langle t_i, d_i, p_i\rangle\rangle$, where t_i is tonality data, d_i is deviation data, and p_i is pitch data, in the *i*-th section of the music. We define the metric space of music by three axes: tonality, pitch, and major/minor. These three elements are significant factors affecting to the impression of music.

Tonality is the structure of music, which is composed of sequential musical notes. There are 24 kinds of tonality consisting of 12 major tonalities and 12 minor tonalities. Tonality changes with time in music, and this causes changes in the impression of the music. In music theory, there is the circle of fifths, which defines the distance or similarity between each pair of the 24 tonalities. Each tonality can be represented by an angular value on the circle of fifths. The system processes this angular value into a tonality scalar *t*, which is a value between 0 and 1. Thus, the system converts the distance measured in hue's angle into the distance measured in tonality's angle.

Major/minor refers to a deviation of tonality within a music section. The system calculates the deviation of tonality in a music section, and converts the deviation value into the major/minor scalar *d* that is a value between 0 and 1, which represent the maximum minor deviation and the maximum major deviation, respectively.

Pitch is a value of pitch in a musical score. The system calculates the average of the pitches in a music section, and converts the average value into the pitch scalar *p* that is a value between 0 and 1, which represent the lowest pitch and the highest pitch, respectively.

B. Primitive Functions

For the relevance computation, the proposed system provides a cross-media relevance calculation function and six distance functions. The system calculates how the media data change with time by applying distance functions to each feature extracted from the media data. Then, the system compares the set of distance values to calculate the relevance of the music to the image query. The following three functions are the distance functions for image query:

- The distance in hue between the *i*-th image and the (*i*+1)-th image is $\Delta_{hi} := |h_i - h_{i+1}|$, where *h* is the hue angle in the HSV cone.
- The distance in saturation between the *i*-th image and the (*i*+1)-th image is $\Delta_{si} := |s_i - s_{i+1}|$, where *s* is the saturation coordinate in the HSV cone.
- The distance in value between the *i*-th image and the (*i*+1)-th image is $\Delta_{vi} := |v_i - v_{i+1}|$, where *v* is the value coordinate in the HSV cone.

The following three functions are the distance functions for music:

- The distance in tonality between the *i*-th section and the (*i*+1)-th section is $\Delta_{ti} := |t_i - t_{i+1}|$, where *t* is the tonality angle in the circle of fifths.
- The distance in tonality deviation between *i*-th section and the (*i*+1)-th section is $\Delta_{di} := |d_i - d_{i+1}|$, where *d* is the deviation in tonality.

 ISBN: 978-1-61208-280-6

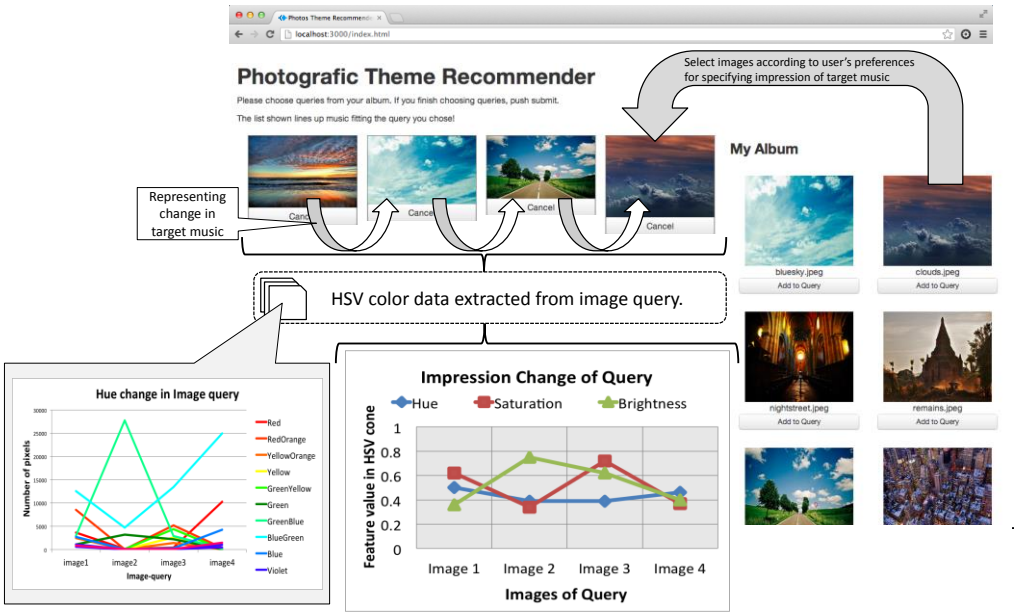

Figure 3. Prototype Implementation that Employs Modern HTML5 Technologies to Edit Image Queries Representing Changes of Mood

- The distance in pitch between the *i*-th section and the (i+1)-th section is $\Delta_{pi} := |p_i - p_{i+1}|$, where *p* is the pitch.

The system provides a fundamental function to calculate the relevance of the music to the query. The function is defined as follows: $(a, b) \rightarrow 1 - |a - b|$, where *a* and *b* form a pair of distance changes according to the dual-metrics relation. The system calculates a correlation value for each pair of metrics by using this function. Moreover, the relevance of the music to the image query is represented as follows:

$$\gamma(\Delta q, \Delta m) := \frac{\sum_{i=1}^{n} \frac{s(\Delta_{hi}, \Delta_{ti}) + s(\Delta_{hi}, \Delta_{ti}) + s(\Delta_{hi}, \Delta_{ti})}{3}}{n} \quad (1)$$

where *n* is the number of images inputted as image-query, as well as the number of divided music sections.

V. Prototype Implementation

We have implemented a prototype of the proposed system. The prototype system is implemented using HTML5 Canvas, jQuery UI, and Backbone.js as shown in Figure 3. The system implemented consists of three modules: the query editor, the feature conversion module, and the music retrieval engine. The query editor is the main user interface. Users can edit a query by selecting four images and revising the order of the selected images. The feature conversion module generates a query by converting the feature of the inputted image sequence into changes of sentiment in the target music. The music retrieval engine calculates the relevance of candidate music to the converted query.

VI. Conclusion Remarks

We have proposed a cross-media retrieval system for music. The system not only analyzes the changes of mood in candidate music and the sequence of images but also calculates the relevance of the music to the images by using a specially designed metric space for change calculation. As future work, we plan to develop a social-network-based query recommendation by this approach.

References

[1] M. Goto and K. Hirata, "Recent studies on music information processing," Acoust. Sci. Technol., vol. 25, no. 6, 2004, pp. 419–425.

[2] D. Temperley, Music and Probability, MIT Press, 2007.

[3] C. L. Krumhansl, Cognitive Foundations of Musical Pitch, Oxford Univ. Press, 1990.

[4] R. Typke, F. Wiering, and R. Veltkamp, "A Survey of Music Information Retrieval Systems," Proc. ISMIR 2005, Univ. of London, 2005, pp. 153–160.

[5] Y. Hijikata, K. Iwahama, and S. Nishida, "Content-based Music Filtering System with Editable User Profile," Proc. ACM SAC'06, ACM Press, 2006, pp. 1050–1057.

[6] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by Humming: Musical Information Retrieval in an Audio Database," Proc. 3rd ACM Conf. Multimedia (Multimedia '95), ACM Press, 1995, pp. 231–236.

[7] F. F. Kuo and M. K. Shan, "Looking for New, Not Known Music Only: Music Retrieval by Melody Style," Proc. 4th ACM/IEEE-CS Joint Conf. Digital Libraries, (JCDL '04), ACM Press, 2004, pp. 243–251.

[8] P. Ciuha, B. Klemenc and F. solina, "Visualization of concurrent tones in music with colours" Proceedings of the ACM MM '10 2010, pp. 1677-1680.

[9] S. Craig, "Harmonic Visualizations of Tonal Music," Proc. Int. Computer Music Conf., 2001, pp. 423–430.

[10] K. Peacock, "Synesthetic perception: Alexander Scriabin's color hearing," Music Percep. vol. 2, no. 4, 1985, pp. 483–506.

[11] S. Imai, S. Kurabayashi, and Y. Kiyoki, "A Music Database System with Content Analysis and Visualization Mechanisms," Proc. IASTED DIMS 2008, pp. 455–460.

 ISBN: 978-1-61208-280-6