Affective Music Recommendation System using Input Images

Shoto SASAKI* Tatsunori HIRAI Hayato OHYA Shigeo MORISHIMA †

Waseda University / JST CREST

1. Introduction

Music that matches our current mood can create a deep impression, which we usually want to enjoy when we listen to music. However, we do not know which music best matches our present mood. We have to listen to each song, searching for music that matches our mood. As it is difficult to select music manually, we need a recommendation system that can operate affectively. Most recommendation methods, such as collaborative filtering or content similarity, do not target a specific mood. In addition, there may be no word exactly specifying the mood. Therefore, textual retrieval is not effective. In this paper, we assume that there exists a relationship between our mood and images because visual information affects our mood when we listen to music. We now present an affective music recommendation system using an input image without textual information.

Our system matches an input image with music using an emotional plane. Russell [1980] proposed the valence/arousal model of affect. Valence refers to positive–negative associations with affective phenomena, whereas arousal refers to energetic–calm associations. All affective stimuli can thereby be defined as a combination of these two independent dimensions. Thus, it is reasonable to use this emotional plane (V-A plane) for matching music and an image.

2. Proposed System

The system consists of three parts. Figure 1 outlines the system.

The first part involves setting the image on the V-A plane using color and texture features. Valdez et al. [1994] demonstrated by means of equations the relationship of saturation S, brightness B with valence V_1, arousal A_1. These equations are given below.

$$V_1 = 0.22S + 0.69B \quad (1)$$

$$A_1 = 0.60S - 0.31B \quad (2)$$

Meanwhile, we can establish associations of coarseness C, direction D with valence V_2, arousal A_2 using canonical component analysis as follow.

$$V_2 = 4.57(D - 0.41) + 3.95 \quad (3)$$

$$A_2 = -0.29(C - 44.93) + 4.26 \quad (4)$$

The learning data is the International Affective Picture System (IAPS), which is a standard affective image set used in psychology. We calculate V-A values using these correlations. In addition, the system integrates (V_1, A_1) with (V_2, A_2) on the basis of distances from the origin of the V-A plane. This is based on impressions being stronger toward the edge of the V-A plane than at the origin.

The second part involves setting the user's music onto the V-A plane. Eerola et al. [2009] have shown using principal components analysis (PCA) that acoustic features can be related to valence and arousal. In this paper, we use the RWC Music database, which has equally populated sets of music categories. Coefficients are computed by PCA with the acoustic features of the RWC Music database. We then multiply these coefficients by the acoustic features of the user's music, thus obtaining V-A values for that music.

The third part is musical recommendation. The system calculates the Euclidean distances between an input image and the user's music on the V-A plane. The system then arranges the distances in ascending order, thereby creating a recommended playlist.

* e-mail: joudanjanai-ss@akane.waseda.jp

† e-mail: shigeo@waseda.jp

SIGGRAPH 2013, July 21 – 25, 2013, Anaheim, California.

ACM 978-1-4503-2261-4/13/07

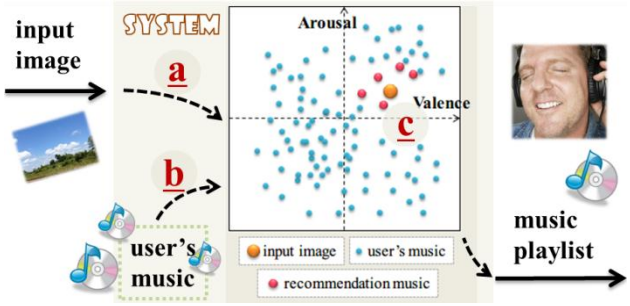

Figure 1: *Outline of Our System.*

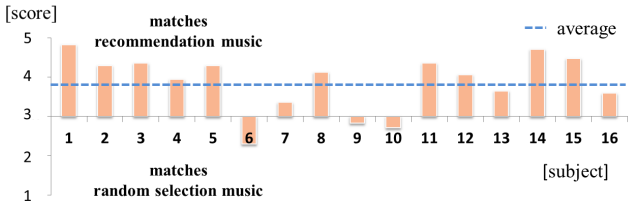

Figure 2: *Result of Evaluation Experiment.*

3. Evaluation

Our proposed method was assessed using subjective evaluation by the random selection method. Sixteen questions were asked to a group of 17 male and female examinees, all in their twenties. The subjects evaluated at five levels on which music is better matched with the image on display. A score of 5 indicated that recommended music matched the image on display. A score of 1 indicated that randomly selected music matched the image on display.

Figure 2 shows the scores of this experiment. The average score is 3.89. Therefore, our proposal technique is confirmed to be effective.

4. Discussion and Conclusion

The results show that questions 6, 9, and 10 have not obtained good scores with the proposed method, particularly question 6, which has the lowest score. The system set question 6 at the origin of the V-A plane, indicating neutral emotion. Hence, images or music in this area tend to affect individuals differently. Therefore, an interactive system is necessary to deal with individuality. For instance, if a user pushes a skip button, the system can learn the user's preferences and make subsequent recommendations by taking them into account.

In this paper, we have presented an affective music recommendation method based on an input image using an emotional plane without textual information. Further work on adding an interactive system is being considered. Moreover, the system can be extended to video input to correspond with change of scenery.

References

RUSSELL, J., 1980. A Circumplex Model of Affect, *Journal of Personality and Social Psychology* 1980, pp.1161-1178

VALDEZ, P., AND MEHRABIAN, A. 1994. Effects of Color on Emotions, *Journal of Experimental Psychology* 1994, pp.394-409

EEROLA, T., LARTILLOT, O., AND TOIVIAINEN, P. 2009. Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models, Proc. *International Society for Music Information Retrieval Conference* 2009, pp.621-626