Automatic Music Tempo and Timbre Change Based on Pitch Detection and Image Analysis

Yongkag Kim, Sangbae Jeong, Minsoo Hahn

Digital Media Lab, Information and Communications University
517-10, Dogok-dong, Gangnam-gu, Seoul, Korea
{ykkim06, sangbae, mshahn}@icu.ac.kr

Abstract

This paper proposes a method for automatically changing musical characteristics by analyzing image features. That is, it presents an automatic music tempo and timbre changing system through piano pitch detection and image analysis. To begin with, the piano's onset is detected, and we extract the pitch based on the onset detection. Next, we analyze the features of the image, such as hue, intensity and roughness. Finally, the music is automatically transformed by the image's features. And we also investigated the correlation between the image and the music by testing the people's preferences.

Keywords: Music, Color, Onset Detection, Pitch Detection

1. Introduction

We usually perceive the world through the visual and auditory information among many senses. People are able to perceive whether the music is sad or delightful while listening to the music, and also perceive whether the color is warm or cold while seeing the image. We usually appreciate paintings in the gallery and surf the internet with background music. And many people post photographs or image file on the web site, such as homepages or blogs in order to represent their daily life or personality. They might also link background music to express their moods or emotional states effectively. The emotion felt from the image goes well with the emotion felt from the music when they uploaded the image and link the music. However, if other images are updated without changing the background music, they sometimes do not go well with. Therefore, we considered about automatic music changing system based on the image analysis. If we find the features of the music, such as tempo and pitch and the features of the image, such as intensity, hue and texture, we can change the musical characteristics by analyzing the image features. That is, we can automatically change the music's mood based on the image's features. And it might make the music harmonious with the image.

There have been many research works on the pitch detection and the relationship between emotion and color.[Eun Yi Kim et al.(2005), Jiajun Zhu et al.(2005), Noriko Nagata(2005), S. Dixson(2001)] Such research works have been focused on specific part. However, we focus on the relationship between the image and the music based on pitch detection and automatic music changing system. In section 2, we describe the system itself: music and image analysis, and transform between image and music. And we present the experimental results about testing preference of the transform in section 3. The conclusion section contains a discussion of the results, and possible future works.

2. Feature Extraction and Transform

The system is composed of three parts: music analysis, image analysis and transform between the features. First part is the music analysis, such as onset and tempo detection. Second part is the analysis of images, such as intensity, hue and texture. Third part is a music synthesis by using the differences of image features. The mapping rule is that the tempo is transformed by the hue, and the envelope contour is transformed by the intensity, and the amount of high frequency is transformed by the texture.

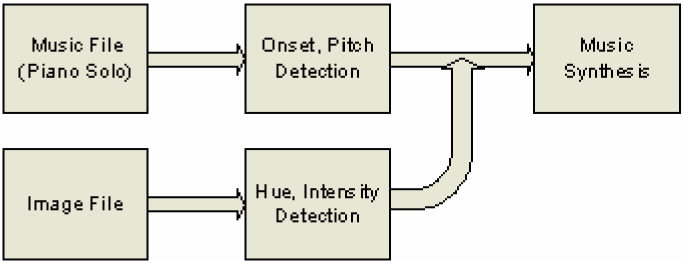


Figure 1. *Overview of the system*

2.1 Feature Extraction from the Music

We simply limited the music as the piano solo and constant tempo music. We detect onset, pitch, and tempo of the music. And they are important characteristics to transform the music automatically.

Onset detection. In this section, we describe the onset detection of the music in the time and frequency domain. The audio file is signal to a 44100Hz sampling rate, 16-bit and mono. First, the signal is divided 10ms interval, and the differences between two adjacent values are obtained, and they are summed. And this process is repeated through the whole signal. Then, we obtain local maximum that is bigger than threshold and the biggest value among the interval. And we decide the local maximum as an onset in the signal. Then, this onset is applied to the tempo and pitch detection.

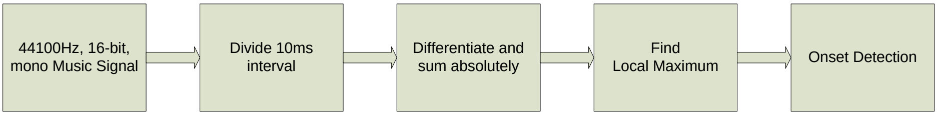

44100Hz, 16-bit, mono Music Signal
Divide 10ms interval
Differentiate and sum absolutely
Find Local Maximum
Onset Detection

Figure 2. *Overview of the onset detection*

Pitch detection. Next, we analyze the frequency by 16384 point FFT(Fast Fourier Transform) to satisfy the piano frequency resolution. If the signal between onsets is less than 16384 points, zeros are padded. Then, FFT is done in the interval after hamming windowing. Among the result of the FFT, we choose the value bigger than the threshold as a pitch. Strictly, this is not a pitch detection of piano, but we focused on the mapping of sound and image and it also helps us to avoid a misdetection of the pitch. There are many values that are bigger than the threshold in one onset, because many harmonics exist. Therefore, we can generate the sound by summing various sinusoidal signals that has a lot of pitch values.

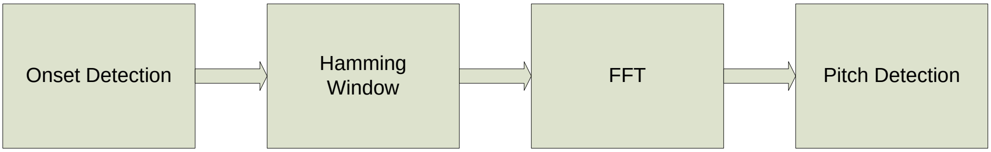


Figure 3. *Overview of the pitch detection*

Tempo detection. We detect the tempo of the music based on onset detection. Figure 4 represents the distribution of the time interval between onsets. We decide the tempo of the music by detecting the peak values. Most of the onsets' intervals are located near the 20 and 40. Also, small amount of the interval are near 60 and 80.

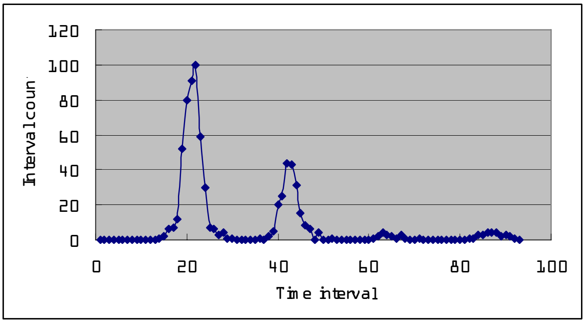


Figure 4. *Distribution of the time interval between onsets*

2.2 Feature Extraction from Image

Hue. Hue values are obtained from the HSI (hue, saturation, intensity) color model. We calculate the average of hue values in each image file. An angle of 0° from the red axis designates 0 hue, and the hue increase counterclockwise from there.

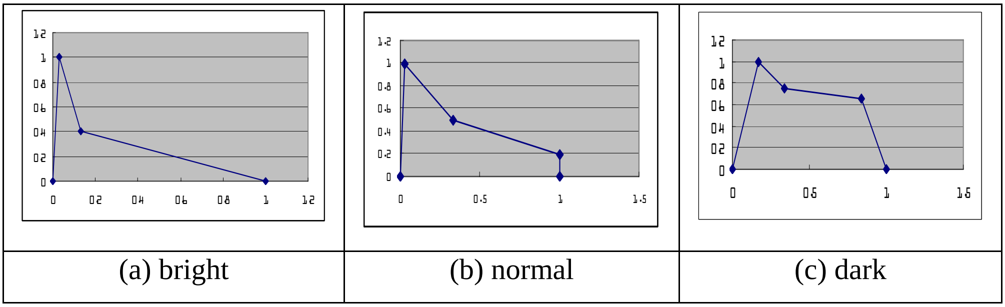
Intensity. Intensity average values are also obtained from the HSI color model. Intensity values are in the range [0, 255]. The image which has a lot of bright parts has a high intensity value, on the other hand a dark image has a low intensity value.

Texture. We transform the color image to intensity image, and find edges using the Sobel approximation to the derivative. Rough image has a lot of edges, and soft image has small amount of edges. After each images' edges are found, it is normalized by the size of the image file.

2.3 Transform between Music and Image

We extract the image's features: intensity, hue and texture. And we apply the features to change the musical characteristics. First, we obtain RGB value of the image and transform to HSI color model. Then, we can obtain hue and intensity value.[Rafael C. Gonzalez et al. (2002)] And the differences of features between images are used to change the musical characteristics.

Hue and tempo. We obtain average hue value from HSI value of the image, and transform the hue value to tempo of the music. Warm color such as yellow-red transforms the tempo slow and cold color such as blue transforms the tempo fast. People usually think that the warmest color is yellow-red and the coldest color is blue. And we linearly transform difference of hue values to the tempo of the music.



(a) bright (b) normal (c) dark

Figure 5. *ADSR envelopes*

Intensity and ADSR(Attack-Decay-Sustain-Release) envelope. We obtain average intensity from the HSI color model of the image. Intensity values are in the range [0, 255], and the envelopes of sounds are like the figure 5. We linearly transform the ADSR envelope based on the intensity. Figure 5-(b)'s envelope represents normal ADSR envelope. The envelop sounds like an ordinary piano tone. The graph of the figure 5-(a) represents the bright image. It has a small amount of sounds comparing with the normal envelope. It has fast decay and falls to zero after fast decay, so it is heard brightly and slightly like marimba. Figure 5-(c) represents a dark image and it has a large amount of sound comparing with the normal envelope. It has a slow attack, small decay, long sustain and steep release. The slow attack and long sustain at a high value make the sound abundant and heavy like trumpet. We transform the brightness of the image to ADSR envelope. The linear interpolation between the each envelopes creates a morph from one instrument to the other, and it makes us perceive the music differently.[R. Boulanger, (2000)]

Texture and timbre. Also, we can obtain the image texture by detecting the edges using the sobel approximation. And we consider the rough image as high value of the sobel operation, and soft image as a low value. Then, we linearly transform the rough image to high frequency emphasized sound, and soft image to low frequency emphasized sound.

2.4 Music Synthesis

We obtained the musical characteristics, such as tempo, pitch, intervals of each pitches, and amplitude. So, we could easily generate the music with the CSound, software sound synthesizer. We already have the CSound's parameters, such as start time, duration, amplitude, and pitch of the signal. We shape the envelope, and assign the tempo, and generate each sinusoidal signal with the parameters. Firgure 6. represents the simplified Csound OpCodes diagram. We shape the envelope by the time duration and parameters by intensity. And the output is multiplied by amplitude, and generate the out signal with pitch.

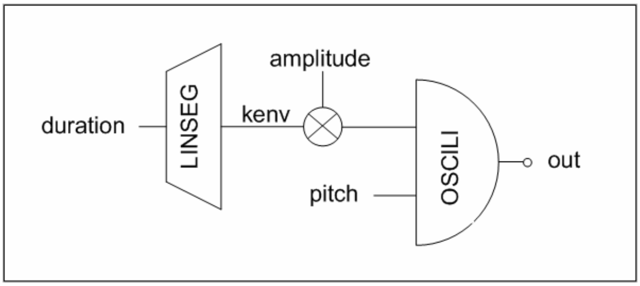


Figure 6. *Diagram of CSound OpCodes*

3. Experiments and Results

This experiment tried to verify whether the mapping rule described section 2 is effective to general people. Thirty subjects of men and women in 20s and 30s are selected to perform the following experiments. First, we test the images that have typical hue, intensity, and roughness. And we investigate the correlation between the image and the music. Also, we applied this test to the photographs and landscape paintings.

3.1 Test images

Hue test images. We chose warm color bar around orange color. Figure 7-(a) is the test images. Upper image is composed of three color, and the color is located in the 0°, 30°, 60° in the HSI color model. And we chose cold color bar around blue color. Figure 7-(a)'s lower image is composed of three color, and the color is located in the 210°, 240°, 270° degree in the HSI color model.

Intensity test images. We also chose bright color bar having high intensity. Figure 7-(b) is the test images. Upper image is composed of three degree of gray color, and the intensity is the 200, 220, 240 in the HSI model. And dark color bar has low intensity value. Figure 7-(b)'s lower image is composed of three degree of gray color, and the intensity value is the 0, 20, 40 in the HSI model.

Texture test images. We chose two kinds of typical images that represent roughness and softness. Figure 7-(c)'s upper image represents a typical soft image. And lower image represents a typical rough image.

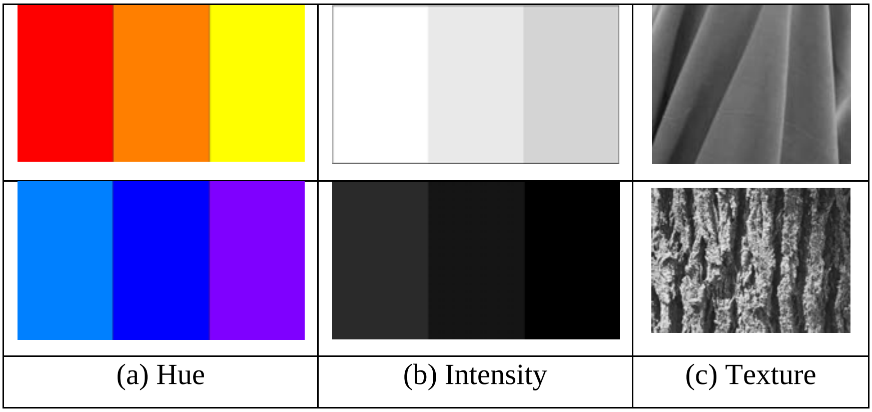

(a) Hue (b) Intensity (c) Texture

Figure 7. *Test image*

3.2 Results

After we transformed the musical characteristics based on the mapping rule, we let the subjects see the two images and let them hear two music. And we asked them to choose each music and image that goes well. The results of this experiment are summarized in Table 1. As we can see in the table, the results show that there are strong correlations on roughness and brightness between image and music. And there is also high correlation on warmth although it is less than other two features. Some subjects commented that they simultaneously felt energetic as well as warm while seeing the test images, and they associated the energy with a fast tempo.

We also applied this mapping rule on photographs and landscape paintings that have typical characteristics of brightness, roughness and warmth. But we cannot get distinct correlation like color bar test, because there are many factors that affect people's emotions. Some subjects set high value of the object such as tree, sun, and mountain rather than color and texture.

Table 1. *Correlation between image and music*

Image	Music	Correlation (%)
Rough	High pass filtered	90.0
Soft	Low pass filtered	
Warm	Slow	73.3
Cold	Fast	
Bright	Light tone	93.3
Dark	Heavy tone	

4. Conclusion

We have presented not only the feature detection of the image and the music but also transform between them. People usually have various kinds of emotions while appreciating the music and the painting, and there are correlations between them based on the experimental results. But it is difficult to apply to a photograph or a painting, because there are many factors consisting of the painting, such as color, objects and so on. But this paper shows the potential to make the music harmonious with the image, and to enhance the moods felt from seeing images by changing the background music's characteristics automatically. However, the current system is limited to the piano solo music and three features of the image and the music. There are many music that is played with many instruments, and images also have many other features as well as color or texture. Future work will investigate the correlations between other features of music and image, and transform the musical characteristics by the image's features.

Acknowledgement

This research was supported by the MIC(Ministry of Information and Communication), Korea, under the Digital Media Lab. support program supervised by the IITA(Institute of Information Technology Assessment).

References

Eun Yi Kim, Soo-jeong Kim, Hyun-jin Koo, Karpjoo Jeong, and Jee-in Kim.(2005), *Emotion-Based Textile Indexing Using Colors and Texture*, LNAI 3613, pp. 1077 – 1080

Jiajun Zhu, Lie Lu.(2005), *Perceptual Visualization of A Music Collection*, ICME 2005, pp. 1058-1061

Noriko Nagata, Daisuke Iwai, Sanae H. Wake, and Seiji Inokuchi.(2005), *Non-verbal Mapping Between Sound and Color-Mapping Derived from Colored Hearing Synesthetes and Its Applications*, LNCS 3711, pp. 401-412.

R. Boulanger, ed. *The Csound Book: perspectives in software synthesis, sound design, signal processing, and programming*, Cambridge, Massachusetts: MIT Press, 2000.

Rafael C. Gonzalez et al. *Digital Image Processing*, Prentice Hall, 2002.

S. Dixson.(2001), *Learning to Detect Onsets of Acoustic Piano Tones*, MOSART Workshop on Current Research Directions in Computer Music.