Emotion-based Impressionism Slideshow with Automatic Music Accompaniment

Cheng-Te Li, Man-Kwan Shan
Department of Computer Science,
National Chengchi University
Taipei, Taiwan
{s9232, mkshan}@cs.nccu.edu.tw

ABSTRACT

In this paper, we propose the emotion-based Impressionism slideshow system with automatic music accompaniment. While conventional image slideshow systems accompany images with music manually, our proposed approach explores the affective content of painting to automatically recommend music based on emotions. This is achieved by association discovery between painting features and emotions, and between emotions and music features respectively. To generate more harmonic Impressionism presentation, a linear arrangement method is proposed based on modified traveling salesman algorithm. Moreover, some animation effects and synchronization issues for affective content of Impressionism fine arts are considered. Experimental result shows our emotion-based accompaniment brings better browsing experience of aesthetics.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information System – *Evaluation/methodology*; J.5 [**Computer Applications**]: Arts and Humanities – *Fine arts*.

General Terms

Algorithms, Design, Human Factors.

Keywords

Slideshow, Music accompaniment, Association discovery.

1. INTRODUCTION

There was a painting-music flash movie, "Vincent", composed by England's singer Don Mclean for the impressionism artist Vincent Van Gogh, distributed over the internet. Along with Mclean's epic-like tone, this movie leads us to the deepest inner world and the soulful sympathy of Van Gogh. The consonance between painting and music is important for fine art slideshow.

Some works have been done on slideshow system for photos. The work proposed by Hua et al. [3], arranges photos by time, scene, or content similarity with incident music. They also proposed a system, Photo2Video [4], to convert photos to motion video with designated music. Photo2Video explores attentive area, topic, and human face of photos and applied "Ken Burns" effects for motion photo clip. Another work proposed by Chen et al. [1] generates music-driven photo slideshow where each frame is tiled by multiple photos.

MM'07, September 23–28, 2007, Augsburg, Bavaria, Germany.

Basically, painting in digital form is one type of photos. Those systems developed for photo slideshow may be employed for painting. However, some characteristics are distinguished painting from photo [2]. Spatial variations of color and number of unique colors in paintings are larger than real scenes. Perceptual edges in photos of real scenes are intensity edges while those of paintings are pure color edges. In particular, paintings have more affective content than photos. Impressionists are more concerned with conveying emotions. Van Gogh had written to his brother, stating "It is emotion, the sincerity of one's feeling for nature that draws us …" Renoir always painted happier aspects of life, expressing joyful emotions and carefree impression. While the incidental music of existing photo slideshow systems is designated manually by users, it is desirable to automatically accompany painting images with music based on the affective content of paintings.

In this paper, we propose the emotion-based approach to generate Impressionism slideshow with automatic music accompaniment. Figure 1 shows an example of slideshow, for nine Impressionism painting, generated by our proposed approach. The paintings are grouped into three clusters based on their affective content. Music is accompanied for each group based the corresponding emotion. For harmonic presentation, an inter-cluster arrangement process is performed. A bridge of music clip between adjacent clusters is created and each display of painting is synchronized with the occurrence of music strong beat.

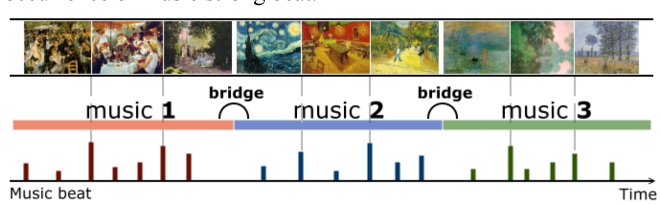

Figure 1. An example of proposed slideshow.

2. SYSTEM OVERVIEW

There are three components in the proposed slideshow system, painting-emotion discovery, music accompaniment and composition, shown in Figure 2. First, for those painting without annotation of emotions, affective features are extracted and the corresponding emotions are discovered. This is achieved by association discovery between painting features and emotions from training data (painting with annotation of emotions). Then, paintings are clustered according to the corresponding emotions. To recommend music for each cluster, the association discovery between emotions and music features is performed to recommend music. Finally, an inter-cluster ordering is generated by arrangement module while each painting is animated by *ROI (region of interest)* determination and motion generation for plentiful spatio-temporal experiences. Then, synchronization is performed by alignment of music strong beat with painting display. Moreover, for the gaps between each painting and each music clip, some strategies are applied to smooth the disharmony.

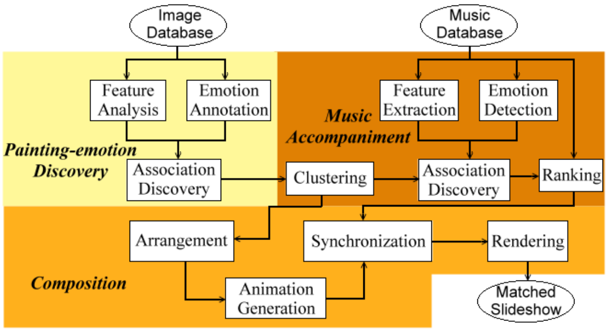

Figure 2. Flowchart of proposed slideshow system.

3. PAINTING-EMOTION DISCOVERY

3.1 Painting Features

Painting is characterized by color, light, texture, line, etc. Among them, the first three affects emotions strongly. Emotions are sensitive to color. For example, Van Gogh's "*The Night Café*" tried to express with red and green the terrible passions of human nature. For color and light, we extract *dominant color*, *color coherence vector (CCV)*, and *color moment*. The extent of coarseness in the surface of painting is decided by texture. There exists relationship between brushwork and texture. For texture, we extract some statistic measures from *gray level co-occurrence matrix (GLCM)*, and compute the *primitive length* of texture.

Dominant Color We calculate the color histogram of a painting. Those colors occur below 10% of area would be omitted. In color description, since the property of impressionists are good at presenting color change by complementary colors, like Van Gogh's "*Wheatfield with Crows*", we utilize the *opponent color theory* formulated by Ewald Hering based on artistic color concepts, in which L means luminance, S means red-green channel, and LM means yellow-blue channel.

Color Coherence Vector Since the color histogram's lack of spatial information, we employ another histogram-based approach for comparing images. The *color coherence* is defined as the degree to which pixels of that color are the members of those large similarly-colored regions, so the pixels would be categorized into coherence or incoherence.

Color Moment Color distribution can be regarded as a probability distribution and characterized by its central moments. It is so-called *color moment*. The color moment computes index regarding only the dominant features instead of entire color distributions. The first three central moments of each color channel are calculated, including average, variance and skewness.

Gray Level Co-occurrence Matrix By considering the spatial relationship of pixels, GLCM measures the frequency of adjacent pixel patterns. It is a 2-d matrix of joint probability $P_{d,r}(i, j)$ between pairs of pixels, separated by a distance, d, in a given direction, r [2]. If the texture is coarse, the pairs of pixels would have similar gray levels. Some features of GLCM are extracted, including energy, contrast, correlation, entropy, and homogeneity.

Primitive Length The pixels of the same gray-level with wide area stand for coarse texture and little area stands for fine texture. A *primitive* is a continuous set of maximum number of pixels in the same direction with the same gray level. Each primitive is defined by its gray level, length and direction. We extract five features: short and long primitive emphasis, gray-level uniformity and primitive length uniformity, and primitive percentage.

3.2 Painting-emotion Association Discovery

The proposed approach accompanies painting with music based on the corresponding emotions. Therefore, we must discover the association between painting features and emotions in advance. The graph-based approach, *Mixed Media Graph (MMG)*, which was originally designed to find correlations across multimedia objects [6], is adopted and modified for the association discovery.

In MMG graph, all the objects and associated attributes are represented as vertices. For objects with n types of attributes, MMG graph will be an (n+1) layered graph with n types of vertices plus one more type of vertices for the objects. There are two types of edges in MMG graph. The object-attribute-value link (*OAV-link*) is the edge between an object vertex and an attribute vertex. The other type, nearest neighbor link (*NN-link*), is the edge between two attribute vertices. An edge is constructed between an attribute vertex and its k nearest neighbor. After the construction of MMG graph, to find the correlations across the media, the mechanism of *random walk with restart* is employed to estimate the affinity of attribute vertices with respect to the query vertices. For the detail computation of affinity, refer to [6].

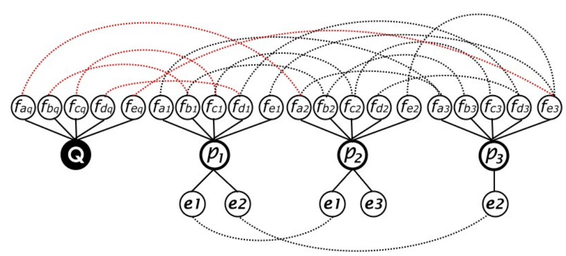

Figure 3. Painting Affinity Graph.

In our approach, we call it *painting affinity graph (PAG)*. In PAG, we create an object vertex for each painting. For each painting object vertex, two types of attribute vertices, emotion and painting feature are created and connected. Five painting feature attribute vertices are created for each painting object vertex, including dominant color, CCV, color moment, GLCM and primitive length. For a painting with m emotions, m emotion attribute vertices are created. NN-link is constructed between two emotion attribute vertices if both are of the same emotion. Nearest neighbor of NN-link between two painting feature attribute vertices is determined by the similarity measure of respect painting feature. Figure 3 shows an example of PAG for three annotated painting. The painting object p_1, with features $\{f_{a1}, f_{b1}, f_{c1}, f_{d1}, f_{e1}\}$, is annotated with emotions $\{e_1, e_2\}$. Given an un-annotated painting Q with painting features $\{f_{aq}, f_{bq}, f_{cq}, f_{dq}, f_{eq}\}$, PAG will find the most likely emotions based on the affinity.

4. MUSIC ACCOMPANIMENT

4.1 Emotion-based Clustering

To recommend music based on the emotions, we have to organize paintings such that those with similar emotions are grouped in a cluster and find out the common emotions of each cluster.

We employ the ROCK algorithm [8] for emotion-based clustering. ROCK is a novel clustering algorithms for data with boolean and categorical attributes.

4.2 Music-emotion Association Discovery

For each cluster of paintings, one music clip is recommended for accompaniment. The basic idea to music recommendation is based on each cluster's common emotions explored by the painting-emotion discovery component. To find the recommended music, association discovery between emotions and music features is performed. The music features used in this work are melody, rhythm and tempo, which are adopted from our previous work [5]. The music-emotion discovery process is based on MMG as well. We call it *music affinity graph*.

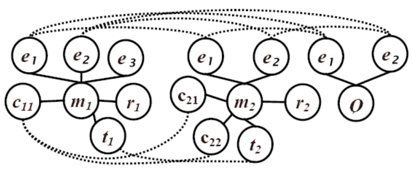

Figure 4. Music Affinity Graph.

Music elements which affect the emotion include melody, rhythm, tempo, mode, key, harmony, dynamics and tone-color. Among these music elements, melody, mode, tempo and rhythm have stronger effects on emotions. Generally speaking, major scale is brighter, happier than minor; rapid tempo is more exciting or tenser than slow tempo. In our previous work [5], we proposed an algorithm to identify the mode of a music piece. We utilized chord as the melody feature and proposed the chord assignment algorithm to take mode and key of melody into consideration. The rhythm feature is derived from beat detection followed by repeating pattern discovery. The feature tempo is calculated from resolution of the music and beat density of the most repetitive pattern. For the detail of music feature extraction, refer to [5].

Figure 4 illustrates a music affinity graph with respect to the query emotions $\{e_1, e_2\}$. Given query emotions, music affinity graph will find the most likely music features based on the affinity. These found music features are then employed to rank the database music and to recommend music for query emotions.

5. COMPOSITION

5.1 Linear Arrangement

In order to generate a harmonic slideshow, linear arrangement is performed for display order of paintings. The basic idea of arrangement is to schedule the display order such that the dissimilarity of affective content between adjacent paintings is as less as possible. The *Earth Mover's Distance (EMD)* [7] is employed to measure the dissimilarity of painting features, since it mimics the human perception of texture similarities.

Recall that paintings with similar emotions are clustered. There are two levels of arrangement, intra-cluster and inter-cluster. The intra-cluster arrangement creates an ordering for paintings within a cluster. Having computed the dissimilarity for each pair of paintings and constructed the corresponding complete graph, we model the intra-cluster arrangement as the *bottleneck Traveling Salesman Problem (TSP)* from developed algorithm, which tries to find the Hamilton cycle with the minimal length of the longest edge. An example is illustrated in Figure 5.

For inter-clustering arrangement, the display order among clusters is scheduled. We compute the dissimilarity between each pair of clusters, which is defined as the dissimilarity between the last painting of one cluster and the first painting of the other cluster. While inter-cluster dissimilarity is asymmetric, we model the inter-cluster arrangement problem as the *Asymmetric Traveling Salesman Problem*. There exist several heuristic algorithms for Bottleneck and Asymmetric Traveling Salesman Problems.

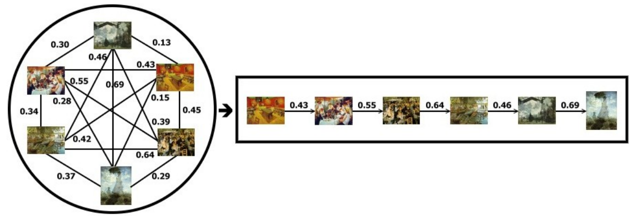

Figure 5. Intra-cluster arrangement.

5.2 Close-up Animation

To highlight the spatial information of affective content of painting, along with the temporal feelings of music, we create the animated effects into each interval of painting display by motion patterns, such as panning and zooming. For these motion patterns, we have to extract some close-up areas in each painting and treat it as key-frames. In this work, the ROI is adopted as the close-up area. The ROIs are determined by face detection and attention model of saliency map. Figure 6 is an example of animation.

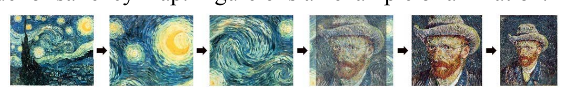

Figure 6. An example of zooming and panning.

5.3 Synchronization and Rendering

The effects of synchronization of temporal structure between paining and accompanied music play a vital role on the affective impression of viewers. The temporal structure congruency emphasizes the simultaneous accents between auditory and visual content. To achieve this, beat information for each music clip is detected and the beginning of each animated clip is aligned to the occurrence of strong beat.

Moreover, though inter-cluster arrangement tries to minimize the distance between adjacent clusters, it is beneficial to reduce the visual gap between the adjacent paintings from two clusters by transition effects. As the dissimilarity between adjacent paintings is below a threshold, the cross-fade effect is applied while beyond this threshold, the cut effect is applied. The duration of transition effect is determined by the strength of music beats.

For the gap between adjacent music clips, we generate musical bridges based on the motives of music clips and the concept of chord progression in Harmony Theory. The process of bridge generation is explained by the illustration in Figure 7. Figure 7 shows three music clips along with corresponding motives and accompanied chords. For example, the motive of the first clip is motive *a* whose pitch contour is a downward curve. The accompanied chords of motive *a* are I and IV. In our approach, musical bridges, shown as dotted box, will be inserted. To generate the bridge, first, the chords of the bridges are generated based on chord progression. Then, melody sequences which harmonized with the generated chords are treated as the candidate motives. Finally, the fitting function of genetic algorithm is employed to generate the most appropriate bridge whose melody contour minimizes the gap between adjacent motives.

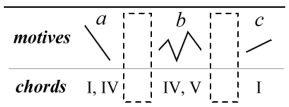

Figure 7. An example of bridge generation.

6. EVALUATIONS

To evaluate the effectiveness of the proposed approach, we evaluate the audiovisual performance by subjective experiments. This evaluation procedure mimics and modifies from that of [1]. We perform experiment on 138 impressionism paintings by nine impressionists, including Claude Monet, Pierre-Auguste Renoir, Vincent Van Gogh, Alfred Sisley, Mary Cassatt, Camille Pissarro, Paul Cezanne, Edgar Degas and Toulouse Lautrec. These images are collected from web sites, http://www.impressionniste.net/ http://www.artchive.com/, and http://fineartfile.idv.tw/. For the accompanied music, we perform experiments on 160 classical piano music from 12 musicians, including Bach, Beethoven, Brahms, Chopin, Debussy, Handel, Mendelssohn, Mozart, Ravel, Schumann and Tchaikovsky. These music objects in MIDI format are collected from web sites http://www.kunstderfuge.com/ and http://www.classicalmusicmidipage.com/. These music objects are segmented into clips by our previous work [5]. The emotion model adopted in this experiment is Russell's circumplex model. Two painting sets used and the ordered music to accompany with are listed in Table 1.

Table 1. Description of painting sets to be evaluated.

Painting set 1	Painting set 2
Number of paintings = 65	Number of paintings = 61
Display length = 5 min. 41 sec.	Display length = 5 min. 17 sec.
1. "*Song without Words, Op.19, No.5*" of Mendelssohn **2.** "*Clair de Lune*" of Debussy **3.** "*Nocturne No.9 in B major, Op.32, No.1*" of Chopin	**1.** "*Moonlight Sonata, No.14, Mv.1*" of Beethoven **2.** "*Scenes from Childhood, Op.15, No. 7*" of Schumann

The slideshows created by ACDSee (http://www.acdsystems.com) and Photo Story (http://www.microsoft.com), are employed in this experiment for comparison. The ACDSee generates one-by-one presentation and is not able to accompany images with music. Photo Story supports some camera effects on photo display, and it is affordable to synchronize between photo and music. However, for those existed music, Photo Story must designate manually.

Table 2. List of five criteria for evaluation.

*Emotionality (**Emo**)*	Do you think it reach emotional coordination?
*Aesthetics (**Aes**)*	How do you feel the presentation of appearance?
*Experience (**Exp**)*	Do you think it reach appreciation aesthetics?
*Atmosphere (**Atm**)*	How do you feel the audiovisual effect?
*Acceptance (**Acc**)*	Do you want to use it to taste fine-art?

We invited 18 persons to participate this experiment. Each person is asked to view the slideshows. The accompanied classical music is of the same length. Then they are asked to give scores from 1 to 10 to express their satisfaction for the questions listed in Table 2, that higher score indicates better satisfaction.

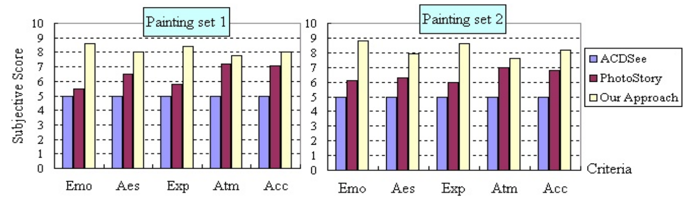

Figure 8. Subjective evaluations for Impressionism paintings.

We set the score of ACDSee to 5 as the baseline. Figure 8 shows the results of subjective evaluations. It is observed that our system gets higher satisfaction than others on average, especially for the criterion of emotionality and experience. This comes from the consideration of the affective content in our proposed approach. However, there is little difference between ours and Photo Story for the criterion of atmosphere. Perhaps this is because the dynamics of beats is not so strong in classical piano music.

7. CONCLUSIONS

In this work, we presented a generic framework to automatically recommend accompanied music for impressionism slideshow based on affective content. The kernel of the proposed approach is the association discovery from painting features to emotions and from emotions to music features. Five features of painting and three of music were extracted. In the composition stage, the TSP algorithm was modified to perform the linear arrangement and the animated effects are applied. Finally, some synchronization issues are considered. Experimental result shows our accompaniment approach gives viewers more pleasant appreciating experience.

8. REFERENCES

[1] Chen, J. C., Chu, W. T., Kuo, J. H., Weng, C. Y. and Wu, J. L. Tiling Slideshow. *In Proc. of ACM Intl. Conference on Multimedia (MM'06),* 2006.

[2] Cutzu, F., Hammoud, R. and Leykin, A. Estimating the Photorealism of Images: Distinguishing Paintings from Photographs. *In Proc. of IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR'03),* 2003.

[3] Hua, X. S., Lu, L. and Zhang H. J. Content Based Photograph Slide Show with Incidental Music. *In Proc. of Intl. Symposium on Circuits and Systems (ISCAS'03),* 2003.

[4] Hua, X. S., Lu, L. and Zhang H. J. Photo2Video - A System for Automatically Converting Photographic Series into Video. *In Proc. of ACM Intl. Conference on Multimedia (MM'04),* 2004.

[5] Kuo, F. F., Chiang, M. F., Shan, M. K. and Lee, S. Y. Emotion-based Music Recommendation by Association Discovery from Film Music, *In Proc. of ACM Intl. Conference on Multimedia (MM'05),* 2005.

[6] Pan, J. Y., Yang, H. J., Faloutsos, C. and Duygulu, P. Automatic Multimedia Cross-modal Correlation Discovery. *In Proc. of Intl. Conference on Knowledge Discovery and Data Mining (KDD'04),* 2004.

[7] Rubner, Y., Tomasi, C. and Guibas L. The Earth Mover's Distance as a Metric for Image Retrieval. *Intl. Journal of Computer Vision, 40(2)*, 2000.

[8] Sudipto, G., Rajeev, R., and Kyuseok, S. ROCK: A Robust Clustering Algorithm for Categorical Attributes. *Information Systems*, *25(5)*, 2000.