ABSTRACT

In this paper, we propose a framework to assess visual complexity of paintings. This framework provides a machine learning scheme for investigating the relationship between human visual complexity perception and low-level image features. Since the global and local characteristics of paintings affect human’s holistic impression and detail perception, we design a set of methods to extract the features that represent the global and local characteristics of paintings. By feature selection, we look into the role that each image feature plays in assessing visual complexity. Then the selected features are combined by a Support Vector Machine for classification. Experimental results indicate that the proposed work can predict the visual complexity perception of paintings with the accuracy of 88.13%, which is highly close to the assessments given by humans. Compared with the conventional measure of complexity, our approach considers human visual perception and performs more efficiently in assessing visual complexity of painting images.

Index Terms— Visual complexity, Perception, Feature extraction, Classification, Machine learning

1. INTRODUCTION

Nowadays, digital technologies and internet make people have more opportunities to appreciate the paintings without going to museums [1]. More and more users select preferable paintings from the internet. If they consider selecting images only by visual feeling (e.g. aesthetic) instead of specific keywords (e.g. flowers), visual complexity has some information of composing the feeling [2, 3]. Hence, presenting an objective index of complexity that fits human feeling to the users is useful.

However, the visual complexity is not directly related to simple objective measures like distribution of spatial frequencies. It is usually affected by visual content (various image features) of paintings, such as the color features, distributions of objects and so on. Hence, in this paper, we aim to propose a framework to assess visual complexity of painting images based on image features, and explore the relationship between human visual complexity and image features. In other words, we attempt to make a computer to sense the visual complexity of a painting as that sensed by the majority of people.

1.1. Visual complexity and related works

In recent decades, visual complexity has become an important and appealing issue. It has been defined in many ways, but there is no uniform definition. Webster’s dictionary (1986) defines a complex object as “an arrangement of parts, so intricate as to be hard to understand or deal with”. According to [4, 5, 6], the definition of complexity should be very close to the numbers of elements in the visual stimulus, their order of placement and the difficulty in description.

Different methods were proposed from psychology to computer science. In the field of psychology, some researchers mainly investigated the factors that affect human visual complexity perception [7, 8, 9]. In the field of computer science, various methods were proposed to measure the complexity, such as, information theory [10, 11], the pattern measure [12], and the fuzzy theory [13]. In [14], Donderi found a correlation between subjective estimations of visual complexity and the size of compressed digital image files.

1.2. Hypothesis

The methods mentioned above evaluate the complexity of images mainly on the basis of information theory and the fuzzy theory, regardless of human visual impression of complexity. In this paper, we attempt to compute the complexity of paintings from the point of human visual perception. Consequently, we assume that the perception of visual complexity is influenced by various characteristics (image features) intuitively perceived by humans.

In order to verify this assumption, we design a group of methods to extract image features that represent both the global and local characteristics of paintings. Inspiration for these features is from a questionnaire survey we conducted to identify the factors that affect human’s visual complexity assessment. By feature selection, we analyze the role that each image feature plays in assessing visual complexity. Then the selected features are combined by a Support Vector Machine for predicting the visual complexity of a painting.
more, we extend the proposed work to architecture images and testify its validity.

2. PROPOSED APPROACH

Assessing visual complexity is a highly subjective task. However, there is a natural intuition that a majority of people with the same background may have a global agreement on classifying visual complexity towards certain paintings. In this paper, we overcome this challenge by introducing a machine learning method, aiming to classify the complexity of paintings into three classes: high complexity (HC), middle complexity (MC), and low complexity (LC).

In order to achieve the purpose of this paper, we conduct three steps: 1) Assessing the visual complexity of paintings and identifying the factors that affect human visual complexity by a psychophysical experiment. 2) Extracting a series of features to globally and locally represent these factors in a painting. 3) Using a machine learning method to build the relationship between the visual complexity perceived by humans and the features extracted from the paintings.

2.1. Subjective assessment of complexity

2.1.1. Experiment setups and procedure

Fifty digital images from the dataset of PaintingDb [15] are used in this experiment. All images were resized to the same height (300px) and randomly displayed on a 46-inch plasma display one by one. Thirty two respondents comes from Hiroshima University participated in the experiment. None of the respondents in this experiment is in major of art or art-related. Their ages ranged from 21 to 30 years old. All respondents had normal or corrected-to-normal vision. The respondents were required to sit 2m from the screen.

The experiment procedure includes two parts. Part I is complexity-rating. Part II is a questionnaire. After the brief introduction of the experiment, Part I was done before Part II.

In Part I, all images were displayed twice. On the first display, the respondent was required to view all images one by one with no time constraint. On the second display, the respondent was asked to score complexity on a 7-point Likert scale according to their perception. The 7-point Likert scale ranged from 1 (very simple) to 7 (very complex). In Part II, we provided a questionnaire with the list of possible factors that affect complexity assessment (the options are listed in the left column of Table 1). The respondent was asked to select the factors which are important for them to assess the complexity of a painting.

2.1.2. Results

The subjective complexity assessments of 50 paintings were obtained in Part I from 32 respondents. We defined “3” and “5” as the threshold for labeling images as “LC”, “MC”, and “HC”. A painting image is marked as LC if its score is lower than 3. A painting image is marked as HC if its score is equal to or greater than 5. And a painting image in-between is marked as MC. Figure 1 shows some examples marked as LC, MC, and HC.

Answers from Part II can be ranked according to the frequency of options mentioned by the respondents (shown in Table 1). It is obvious that the top three frequently mentioned options are “Distribution of compositions”, “Colors”, and “Contents”. Hence, we identify that these three factors importantly affect respondents’ assessment of complexity.

2.2. Feature extractions

Synthesizing the results in the experiment and the common sense in art or the intuition, we extract a series of features to represent the above three factors, and then evaluate whether these features are useful or not.

All these features are separated into two categories: global features and local features. Global features refer to the characteristics of the first impression when human beings see a painting. On the other hand, the local features reflect the regional information of this painting.

2.2.1. Global features

As we all know, no matter in art or in the daily life, it turns out that when viewing something, people firstly get a holistic impression of it and then go into segments and details [16]. Each global feature is shown and explained as follows.

Colors
$f_{1-4}$: Color complexity in four levels of image pyramid are extracted. Colors are the basic elements of a painting. If the colors in a painting are complex, the painting is also of complexity. To measure the complexity of the colors in a painting, we employed the method of Color Complexity Measure (CCM) [17]. Some features hidden in this resolution are extracted in another resolution. Therefore, we applied Gaussian pyramid on the painting images and calculated four features respectively on the four levels of pyramid.

$f_{5-7}$: The average hue ($f_5$), saturation ($f_6$) and brightness ($f_7$) of a painting based on HSL (hue, saturation, and lightness) color space. Hue is the most obvious characteristic of a color [18]. Saturation measures the intensity of a color. Lightness reflects the tone of a painting. These features virtually affect human visual complexity perception of paintings.

Contents

$f_8$: Symmetry plays a relevant role in perception problems [19]. It is an interesting property in detecting the points on interest. The more the points of interest in an image, the more complex the image is perceived to be [13]. We employed the method in [13] to calculate $f_8$.

$f_9$: Edge density of a painting. The edge density can be determined by the ratio between the pixel number of the extracted edges and the pixel numbers of the whole image.

2.2.2. Local features

Local features represent the detail information of the paintings, which may be more attractive for the viewers’ deep viewing. To extract the local features in the painting, we firstly segment the image into parts, and then analyze the characteristics in segments.

Image segmentation

In this paper, an initial segment is firstly required to partition the image into small regions for merging. By comparison with the method of watershed, we chose mean shift for initial segment because it creates less over segmentation. We used a free software, EDISON System [20], to obtain the initial segmentation map. After the initial segmentation, the image is subdivided into many small regions.

In human visual perception, some regions with similar color or spatially adjacency should be merged into one region. Hence we need to represent these regions by some feature descriptors and define a rule for region merging. In this paper, we employed the method of color histogram similarity [21] to calculate the similarity between two adjacent regions. Each color histogram is quantized into 16 levels and then total 4096 bins in each region. The color histogram similarity $\rho$ is calculated between two adjacent regions (region $P$ and region $Q$) using the Bhattacharyya coefficient.

$$\rho(P, Q) = \sum_{u=1}^{4096} \sqrt{Hist_P^u \ast Hist_Q^u},$$  \hspace{1cm} (1)

where $Hist_P^u$ and $Hist_Q^u$ are the normalized histogram of adjacent regions $P$ and $Q$. $u$ means the $u$th bin of them. The higher the Bhattacharyya coefficient between $P$ and $Q$, the higher the similarity between them is.

After calculating the similarity between two neighbor regions, we used the Region Adjacency Graph (RAG) [22] to store the similarity of the pair of regions. The merging rule is defined as: if the similarity between adjacent regions $P$ and $Q$ is the maximal one among all the similarities, we will merge $P$ and $Q$. In this experiment, we loop two times for merging the adjacent regions and got the final segmentation map.

It can be understood that human vision is sensitive to the large segments in the images. Hence, we extract the features in the first largest segment (FLS) and the second largest segment (SLS).

Colors

$f_{10-15}$: Hue ($f_{10}$), saturation ($f_{11}$) and lightness ($f_{12}$) of the FLS. Hue ($f_{13}$), saturation ($f_{14}$) and lightness ($f_{15}$) of the SLS.

$f_{16-18}$: The contrast of hue ($f_{16}$), saturation ($f_{17}$) and lightness ($f_{18}$) between the FLS and its neighbor segments. The calculations are listed as follows.

$$f_{16} = \max_i |H_{\text{largest}} - H_i|, i \in \Omega_{\text{nei}} \hspace{1cm} (2)$$

$$f_{17} = \max_i |S_{\text{largest}} - S_i|, i \in \Omega_{\text{nei}} \hspace{1cm} (3)$$

$$f_{18} = \max_i |L_{\text{largest}} - L_i|, i \in \Omega_{\text{nei}} \hspace{1cm} (4)$$

where $\Omega_{\text{nei}}$ is the set of the neighbor segments around the FLS. $H_{\text{largest}}$, $S_{\text{largest}}$ and $L_{\text{largest}}$ are hue, saturation and lightness values of the FLS, and $H_i$, $S_i$ and $L_i$ are hue, saturation and lightness values of the $i$th neighbor segment.

Distribution

$f_{19}$: Number of all segments. Usually, the larger the number of all segments in the painting image, the more complex the image is.

$f_{20, 21}$: Areas of the FLS ($f_{20}$) and the SLS ($f_{21}$). The larger the FLS, the more homogenous the region is. And this will create gentle visual perception.

$f_{22, 23}$: Shape complexity of the FLS ($f_{22}$) and the SLS ($f_{23}$). In this paper, we use Perimetric Complexity [23] to measure the shape complexity of the FLS and the SLS.

Totally, we extract 23 features that represent both the global and the local features of the paintings.

2.3. Objective measure of complexity

The extracted features were combined by an SVM for classification. We chose a sigmoid kernel function because it yielded the best performance. An SVM model using a sigmoid kernel function is equivalent to a two-layer, perception neural network. We employed 40*32 data as the training data and the rest 10*32 data as the testing data.
which indicates that these features play pivotal roles in predicting visual complexity. We will explain the meanings of these features from human visual perception and psychology theory.

\( f_1 \) and \( f_2 \) represent the color complexity of first two levels of image pyramids. Color complexity is measured by color variation in the local regions of a painting. The higher the color variation in the local region, the more complex the region is. So the painting is also of complexity.

\( f_8 \) is the feature of points of interest. It represents the content of the paintings. The more points of interests in the painting, more complex the painting is.

\( f_{10} \) and \( f_{16} \) are related to the hue information of the largest segment. In the painting theory, the hue reflects the color of a painting [24]. It represents the colorful keynote of the painting. The more colorful the painting is, the more complex the painting is perceived to be.

We compared the proposed method with the conventional method in [14]. Correlations (Pearson correlation) were calculated between subjective complexity assessments and objective measures of complexity. In Table 3, correlations comparison shows that the proposed method is more efficient in assessing visual complexity of paintings.

Moreover, we testified the validity of our method in assessing complexity in architecture images [25]. We used 122 training samples and 75 testing samples in the classification. The results in Table 4 indicate the validity of the proposed method in the architecture images.

### 4. CONCLUSIONS

In this paper, we proposed a new framework to assess visual complexity of paining images. This framework provides a machine learning scheme for exploring human visual complexity and image features. Compared with the conventional measure of complexity, our work considers human visual perception and performs efficiently in assessing visual complexity of painting images.

We showed a slight extension of our approach to architecture images. In the future work, we will further investigate the validity of our approach in other images and applications (e.g. watermarking capacity estimation).

---

**Table 2.** Confusion matrix of predict classification

<table>
<thead>
<tr>
<th></th>
<th>LC</th>
<th>MC</th>
<th>HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td>63</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>MC</td>
<td>0</td>
<td>30</td>
<td>62</td>
</tr>
<tr>
<td>HC</td>
<td>1</td>
<td>30</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>32</td>
<td>224</td>
</tr>
</tbody>
</table>

**Table 3.** Correlations comparison between subjective complexity and objective measures of complexity

<table>
<thead>
<tr>
<th></th>
<th>Proposed method</th>
<th>Method in [14]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlations</td>
<td>0.8805</td>
<td>0.8341</td>
</tr>
</tbody>
</table>

**Table 4.** Classification accuracies of visual complexity in painting images and architecture images

<table>
<thead>
<tr>
<th></th>
<th>Painting images</th>
<th>Architecture images</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>88.13%</td>
<td>73.33%</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Classification accuracies of 24 feature combinations.

---

2.3.1. Feature selection

In order to obtain the feature combination that yields the best performance, we replicated the experiment by 23 times while reducing one feature for each time. By this way, we can also look into the role that each feature plays. A total of 24 classification accuracies are shown in Fig. 2. In this figure, “None” marked in x axis means that we used all features for classification. The subsequent marks mean the deletion of the current feature in each classification. By analyzing this figure, we conclude that there are two aspects of influences for the classification accuracy (CA). One is that the deletions of some features improve or keep the CA, such as \( f_3, f_{12}, f_{14} \) and \( f_{15} \). Another is that the deletions of some features severely decrease the CA, such as \( f_1, f_2, f_8, f_{10} \) and \( f_{16} \). Since the deletions of \( f_3, f_{12}, f_{14} \) and \( f_{15} \) keep or even improve the CA, we decided to remove these indifferent features. Finally, we employed the selected 19 image features for classification.

2.3.2. Classification performance

The classification performance of the proposed method is shown in a confusion matrix (Table 2). The classification accuracy is 88.13%. We also calculated the Kappa coefficient \( \kappa \), which estimates the agreement between subjective evaluations and predicted results. The \( \kappa \) is equal to 0.8805. According to the interpretation of Kappa, the \( \kappa \) value obtained indicates that the complexity predicted from our method keeps accordance with human visual complexity.

3. DISCUSSION

Figure 2 shows the classification accuracies of different feature combinations. It is obvious that the deletions of \( f_1, f_2, f_8, f_{10}, f_{16} \) severely decrease the classification accuracy, which is equal to 0.8805. According to the interpretation of Kappa, the \( \kappa \) value obtained indicates that the complexity predicted from our method keeps accordance with human visual complexity.

In this paper, we proposed a new framework to assess visual complexity of paining images. This framework provides a machine learning scheme for exploring human visual complexity and image features. Compared with the conventional measure of complexity, our work considers human visual perception and performs efficiently in assessing visual complexity of painting images.

We showed a slight extension of our approach to architecture images. In the future work, we will further investigate the validity of our approach in other images and applications (e.g. watermarking capacity estimation).
5. REFERENCES


