We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

On the Design of a Photo Beauty Measurement Mechanism Based on Image Composition and Machine Learning

Chin-Shyurng Fahn and Meng-Luen Wu

Additional information is available at the end of the chapter

Abstract

In this chapter, we propose a machine learning scheme on how to measure the beauty of a photo. Different from traditional measurements that focus on the quality of captured signals, the beauty of photos is based on high-level concepts from the knowledge of photo aesthetics. Because the concept of beauty is mostly defined by human being, the measurement must contain some knowledge obtained from them. Therefore, our measurement can be realized by a machine learning mechanism, which is trained by collected data from the human. There are several computational aesthetic manners used for building a photo beauty measurement system, including low-level feature extraction, image composition analysis, photo semantics parsing, and classification rule generation. Because the meaning of beauty may vary from different people, the personal preference is also taken into consideration. In this chapter, the performance of two computational aesthetic manners for the perception of beauty is evaluated, which are based on image composition analysis and low-level features to determine whether a photo meets the criterion of a professional photographing via different classifiers. The experimental results manifest that both decision tree and multilayer perceptron-based classifiers attain high accuracy of more than 90% for evaluation.

Keywords: computational aesthetics, photo beauty measurement, image composition, machine learning, decision tree, multilayer perceptron

1. Introduction

The computational aesthetics is a field of research to measure the beauty of photos. There are many benefits for predicting the aesthetics score of a photo, for example, computers can aid to manage a huge amount of photos according to the perception of beauty, and can assist to predict whether a photo will be favorable when it is made public in advertisements. The perception of

the beauty of a photo does not concentrate on the measurement of signal quality only, but also cares about the meaning of the concept of beauty defined by human being. In other words, a sharp and clear photo is not always more beautiful than those are not, but it depends on the contents in the photo that delights people. Usually, photos taken by professional photographers are better than amateurs. Many works focus on determining whether an input photo is taken by professionals [1]. **Figure 1** illustrates an example of two photos taken by a professional and an amateur, respectively. The photo in **Figure 1(a)** is taken by a professional photographer, which has better contrast, color harmony, and sharpness. Besides, the contents of the photo are rather simple. In **Figure 1(b)**, the photo taken by an amateur has less color harmony apparently, which also possesses motion blur. Because computer algorithms can measure the contrast, color harmony, and sharpness, determining whether a photo is taken by a professional is possible.

However, the meaning of beauty is different from people. Essentially, people in different locations or of different ages have different tastes. The tastes of eastern and western people differ obviously, and the tastes of old and young are not the same usually. It reveals that subjectivity does exist in the perception of beauty. In order to deal with the subjectivity, the photo beauty measurement system should be flexible, which can be updated and constructed for different cases. In brief, the measurement system is a classification model that can be trained with respect to different kinds of scenarios.

Currently, with the evolution of computer vision, the topic of computational aesthetics arises. By analyzing the attributes and extracting the features of a photo, computers are able to classify whether a photo is preferred by professional photographers or not. There are many kinds of attributes that can be retrieved from a photo, such as brightness, color contrast, saturation, existence of human faces, animals, sky, and so on. After collecting a huge amount of photos in accordance with their attributes and features, a classification model called classifier can be built through training. The trained classifier can be used to predict the aesthetics score of a photo to distinguish good and bad photos. However, depending on training methods, some of the classifiers are like a black box, where their decision process is not understandable. For example, a multilayer perception (MLP) works mostly on the weights of its synapses connecting with neurons, whose decision process is not readable by humans. The decision tree model is a better choice in this aspect, because every path in the decision tree is a
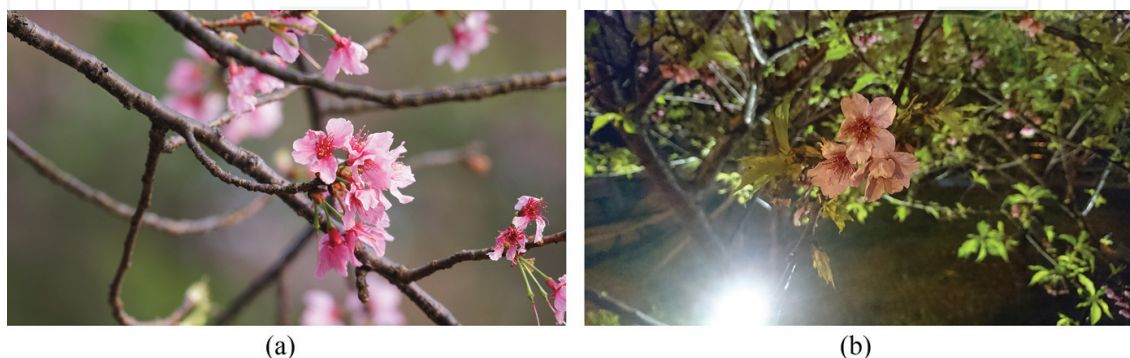

(a)   (b)

Figure 1. Two flower photos for beauty measurement: (a) a photo with high aesthetics score; (b) a photo with low aesthetics score.

readable classification rule. Nevertheless, the accuracy of the decision tree model is sometimes worse than the neural networks or other algorithms that do not produce readable classification rules.

In this chapter, we propose a machine learning scheme to measure the beauty of photos. Two computational aesthetics manners for the perception of beauty are tested; the first is with the aid of low-level features, and the second is resorted to image composition analysis. For such photo beauty measurements, we use two machine learning approaches, which are based on neural networks and decision trees. The neural network model has higher accuracy, while the decision tree produces readable classification rules for humans, which is possible for us to perform photo enhancement according to the rules.

The remainder of this chapter is organized as follows. In Section 2, we list some related works of the perception of beauty for digital photos. In Section 3, we elaborate photo aesthetics with regard to low-level feature extraction. In Section 4, we explain photo semantics that possibly influences the beauty measurement. Section 5 describes image composition analysis used for perceiving the beauty of photos. In Section 6, two machine learning approaches of neural networks and decision trees are presented. In Section 7, we take account of personal preference to solve the subjectivity problem by adjusting the bias of input feature values. Section 8 evaluates two computational aesthetics manners for the perception of beauty according to low-level features and image compositions, respectively. Finally, some conclusions are made in Section 9.

2. Related works

There are some existing systems and methodologies for assessing the beauty of photos. Yeh et al. proposed a personalized photo ranking system to assess the beauty of photos manually with some defined criterions [2]. In the system, users have two options to carry out personalization; one is feature-based, and the other is example-based. The photo ranking system is illustrated in **Figure 2**. In the feature-based option, the system provides a series of feature weights to be adjusted, and the photos will be sorted in the light of their weighted feature
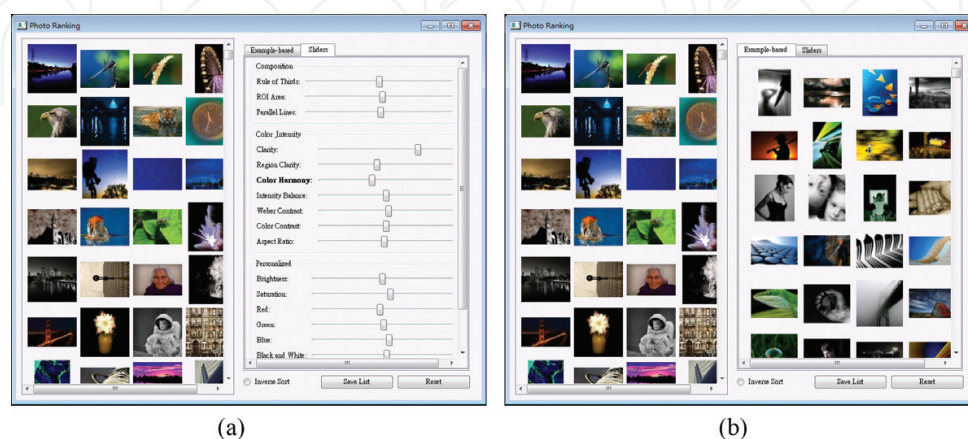

(a)   (b)

Figure 2. Two kinds of personalized photo ranking options: (a) feature-based; (b) example-based.

scores as **Figure 2(a)** shows. In the example-based option, the user selects a number of interested photos, and the system extracts the features of these selected photos to produce weights for features automatically as shown in **Figure 2(b)**. The authors also proposed some new features for photo ranking.

In 2012, an intelligent photographing interface with on-device aesthetic quality assessment system is proposed [3], which makes use of five aesthetics perspectives of photography, such as saturation, color, composition, contrast, and richness. The aesthetic quality assessment system works on a tablet with a camera and runs in real time. **Figure 3** graphically shows the system, where **Figure 3(a)** is the overall rating of features in a photo, while **Figure 3(b)** is a working screenshot of the system.

In consideration of the subjectivity, a digital photo challenge (DPC) platform is established. The platform allows experts to rate a photo at one of 10 aesthetic quality levels, from good to bad. **Figure 4** illustrates some photos in the database for example. In **Figure 4(a)**, the left photo is focused on the flowers successfully, and its theme is harmonic, which makes a comfortable feeling. The color of the right photo is also harmonic and looks comfortable. However, in **Figure 4(b)**, two photos are out of focus, with messy colors and motion blur. Most people would agree that photos in **Figure 4(a)** are better than those of **Figure 4(b)**.
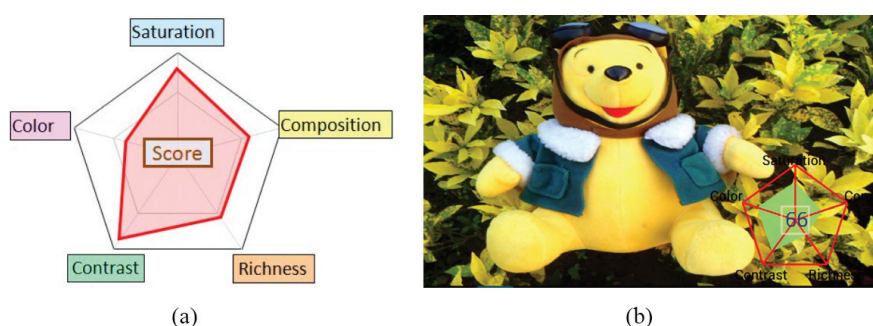

(a)   (b)

Figure 3. An instant aesthetics quality assessment system: (a) five aesthetics perspectives of photography; (b) an assessment example of the system.
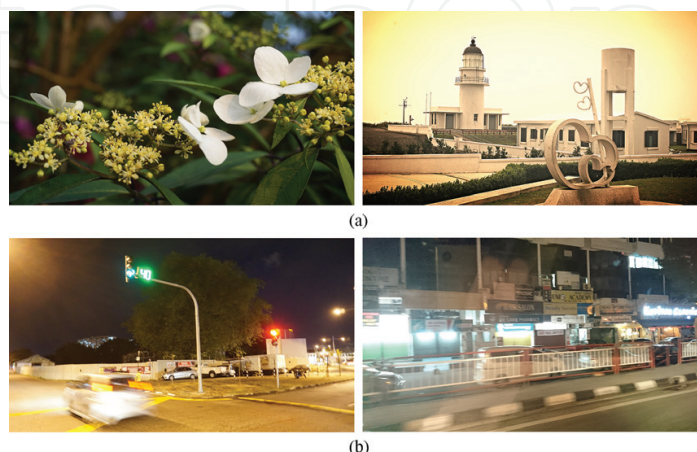

(a)

(b)

Figure 4. Photos ranked by professionals in a DPC platform: (a) high-score photos; (b) low-score photos.

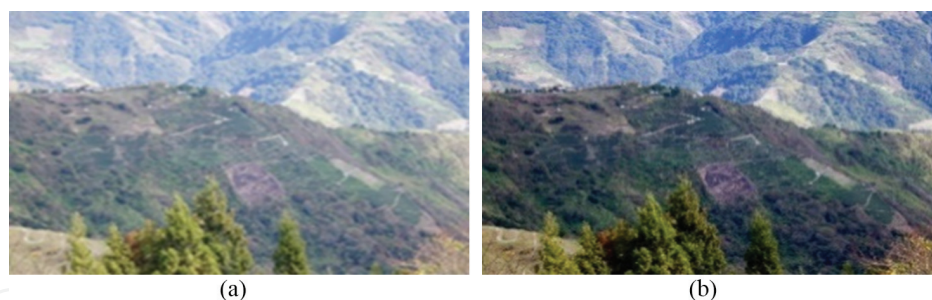(a)   (b)

Figure 5. Photo enhancement by the instruction of a decision tree: (a) before processing; (b) after processing.

In Damon Guy's article [4], he defined photographic aesthetics objectively. He analyzed what elements are used in assessing the beauty of photos, which is governed by the "Principle of photographic art." He found there are 15 important elements including unity, harmony, color, variety, movement, contrast, balance, proportion, pattern, rhythm, geometry, focus, viewpoint, blur, and sharpness. Of these features, the geometry element comprises a photo's shape and composition. The viewpoint element examines whether a person in a photo looks at the camera or not. Some elements are easy to understand and implement; therefore, we will emphasize some photograph features, such as color component, sharpness, brightness, contrast, saturation, color balance, colorfulness, and simplicity in this chapter.

The development of computational aesthetics is also helpful for photo enhancement. In 2016, a photo enhancement method based on computational aesthetics was proposed [5]. In their proposal, a decision tree was produced by virtue of machine learning techniques, and a photo was adjusted to meet the conditions of a favorable contemporary style photo according to the tree, if it is classified as not acceptable. An example of photo enhancement is shown in **Figure 5**, where **Figure 5(a)** is an original photo, and **Figure 5(b)** is an adjusted photo by the proposed method that uses only one instruction to improve the input photo. The example in **Figure 5** just reduced its brightness via their trained decision tree.

3. Low-level feature extraction

Choosing appropriate aesthetic features in photos is essential for distinguishing professional and nonprofessional photos because it helps to predict whether a photo is favorable. Next, we introduce several types of aesthetic features as well as illustrate them. In this section, we focus on low-level feature extraction to measure some elements for the perception of beauty. Such low-level features include color component, sharpness, brightness, contrast, saturation, color balance, colorfulness, and simplicity.

3.1. Color component

The color component feature is acquired from extracting the levels of specific colors in a photo [1]. **Figure 6** shows photos with different major color components.
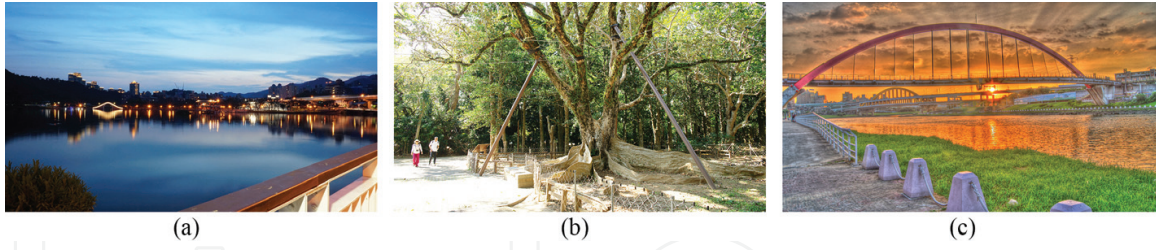
Figure 6. Three photos with different major color components: (a) blue; (b) green; (c) red.

For accuracy, we choose the color component in the CIELab color space to measure. To achieve this, a photo in the RGB color space is necessary to be converted into the CIEXYZ color space first. The conversion matrix is expressed as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

After the photo is transformed into the CIEXYZ color space, it can be then transformed into the CIELab color space by the relation between these two color spaces depicted below

$$L^* = \begin{cases} 116 \times \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, & \frac{Y}{Y_n} > 0.008856 \\ \\ 903.3 \times \frac{Y}{Y_n}, & \text{otherwise} \end{cases} \tag{2}$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]$$

The color component is extracted by the following equation:

$$f_{\text{colorcomponent}} = \sum_{x=1}^{width} \sum_{y=1}^{height} \frac{D(c_l, \text{c}(x, y))}{width \times height} \tag{3}$$

where *width* and *height* are the dimensions of the photo, and $D(c_l, \text{c}(x, y))$ is the Euclidean distance between two colors in the CIELab color space, c_l is the color component we want to extract in the CIELab color space, and $\text{c}(x, y)$ is the color value of coordinate (x, y) in the same color space.

3.2. Sharpness

A blurry photo is almost worse than a sharp photo of the same scene. However, a partially blurred photo is not necessarily unfavorable because the blur may be produced from back-ground defocus using high-end cameras. **Figure 7** shows two photos with different degrees of

Figure 7. Two photos with different degrees of sharpness and blur: (a) a sharp photo; (b) a photo with background defocus.

sharpness and blur in which **Figure 7(a)** has high sharpness whereas **Figure 7(b)** has more blurred regions.

A quality measurement for the sharpness of a photo is stated as follows:

$$f_{\text{sharpness}} = \frac{|\{(u, v)||F(u, v)| > \xi\}|}{width \ \times \ height} \propto \frac{1}{\sigma} \tag{4}$$

$$f_{\text{blur}} \propto \frac{1}{f_{\text{sharpness}}} \tag{5}$$

where $I_{blur} = G_\sigma * I$ is the blurred photo derived through convolving the original photo I with a Gaussian filter G_σ, σ is its standard deviation, and $F(u, v) = FFT (I_{blur}(x, y))$ is the blurred photo transformed into the frequency domain via the fast Fourier transform. Here, ξ is set to 5.

3.3. Brightness

For an input photo, the global brightness can be obtained from various kinds of methods, including software and hardware measurements. In a software method, the global brightness can be calculated by the use of the mean or median of all pixel values. As shown in **Figure 8**, applying two brightness settings to the same scene yields quite different effects for viewers. **Figure 8(a)** is a brighter version, while **Figure 8(b)** is a darker version of the same scene.

Figure 8. Two photos with different degrees of brightness: (a) brighter; (b) darker.

For an input photo, the global brightness can be derived from the following equation:

$$f_{\text{brightness}} = \frac{\sum_{x=1}^{width} \sum_{y=1}^{height} I(x, y)}{width \times height} \tag{6}$$

where $I(x, y)$ is the intensity of a pixel at (x, y).

3.4. Contrast

Color contrast is essential for photo quality measurement because better cameras produce better color contrast. The comparison of different degrees of color contrast is shown in **Figure 9**, where **Figure 9(a)** has both bright areas and dark areas with various colors, while **Figure 9(b)** has only dim white colors.

The color contrast feature is defined as

$$f_{\text{contrast}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (1 - d(i, j)) \frac{D(i, j)}{A_i A_j} \tag{7}$$

where $d(i, j)$ is the spatial distance between the centroids of two segmentations A_i and A_j; $D(i, j)$ is the color distance between the two segmentations in the CIELab color space.

3.5. Saturation

Appealing photos usually have a higher saturation degree. **Figure 10** shows an example for comparison in which **Figure 10(a)** has more vivid colors whereas most pixels in **Figure 10(b)** appear pale and white.

The color saturation feature is defined as

$$f_{\text{saturation}} = \sum_{x=1}^{width} \sum_{y=1}^{height} \frac{s(x, y)}{width \times height} \tag{8}$$

where $s(x, y)$ is the saturation of a pixel in the "hue", "saturation", and "value" (HSV) color space.

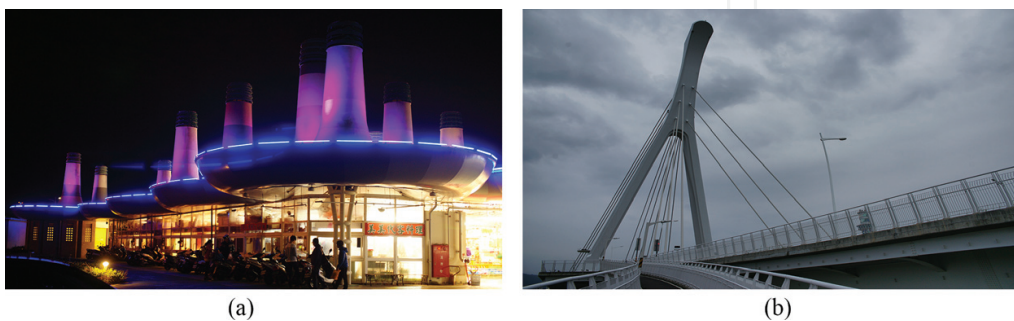(a)   (b)

Figure 9. Two photos with different degrees of contrast: (a) more contrast; (b) less contrast.
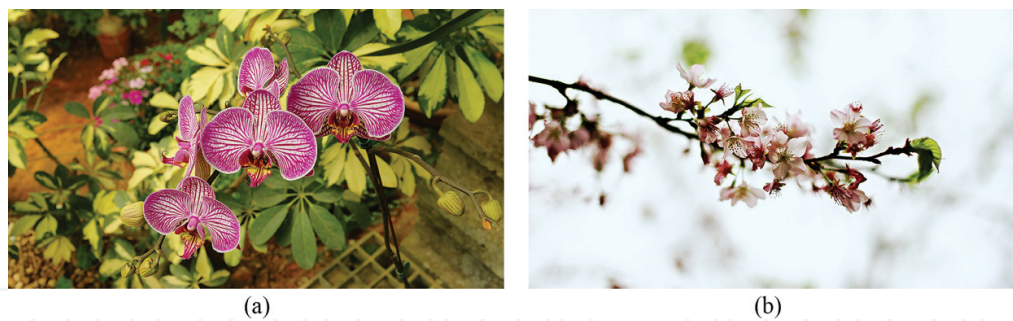
Figure 10. Two photos with different degrees of saturation: (a) more saturated; (b) less saturated.

3.6. Color balance

In the photo aesthetics field, the balance degree of a photo is a good criterion for distinguishing whether a picture is taken by a professional photographer. Professional photographers tend to distribute the color intensity of a photo in a more balanced fashion. A comparison of balanced and unbalanced photos is illustrated in **Figure 11**. In **Figure 11(a)**, the left and right parts of the photo are more balanced, while in **Figure 11(b)**, the photo is less balanced. Usually, a balanced photo has better composition but it is not necessary that unbalanced photos are unfavorable, which is based on the content and the emotion that the photographer wants to express via the combination of varied photo features.

The difference of brightness of the two separated areas can be adopted to obtain this feature. The balance degree of a photo is calculated by

$$f_{\text{balance\_horizontal}} = e^{-(I_{left} - I_{right})^2} \tag{9}$$

where I_{left} and I_{upper} are the average intensities of the left and right parts of the photo, respectively. For the vertical balance feature, the similar equation can be acquired from simply replacing I_{left} with I_{upper} as well as replacing I_{right} with I_{lower}.

3.7. Colorfulness

The colorfulness of a photo is in proportion to the number of nongrayscale pixels in the photo. An achromatic photo, on the other hand, is a grayscale photo. It is not necessary that less

Figure 11. Two photos with different degrees of balance: (a) more balanced; (b) less balanced.

(a)   (b)

Figure 12. Two photos with different degrees of colorfulness: (a) more colorful; (b) more achromatic.

(a)   (b)

Figure 13. Two photos with different degrees of simplicity: (a) simpler; (b) more complex.

colorful photos are low quality because professional photographers sometimes choose elimi-nating colors to express some feelings. As shown in **Figure 12(a)**, the photo is colorful, while the photo in **Figure 12(b)** is achromatic, and they give different feelings. The degree of color-fulness of a photo is defined as the reciprocal of the achromatic feature that is a special color component feature because it comprises rare hue components and it is perceived as a grayscale photo. The achromatic feature can be obtained from

$$f_{achromatic} = \sum_{x=1}^{width} \sum_{y=1}^{height} \frac{|\{\ (x,y)|Ch_R(x,y) = Ch_G(x,y) = Ch_B(x,y)\}|}{width \times height} \tag{10}$$

Then the degree of colorfulness yields

$$f_{colorfulness} = 1/f_{achromatic} \tag{11}$$

3.8. Simplicity

Professional photos are usually possessed of greater simplicity to make the subject appear more attractive. **Figure 13(a)** is simpler in terms of its color distribution whereas the color distribution of **Figure 13(b)** is rather complex.

The simplicity feature is computed from the color distribution of a photo. The formula for the simplicity feature [6] is expressed as

$$f_{\text{simplicity}} = \left(\frac{|\{l|k(c_l) \geq \gamma k_{max}\}|}{4,096} \right) \times 100\% \tag{12}$$

where $k(c_l)$ is the color count for color c_l, k_{max} is the maximum color count, and γ is set to 0.001. In this formula, the number of colors in the photo is reduced to 4,096; that is, the numbers of colors for R, G, and B are all reduced to 16, each of which is represented by 4 bits individually.

4. Photo semantics

A photo with some semantic meanings can be popular even if it lacks some aesthetic elements. For example, in a collection of photos captured by travelers, those with animals or human faces are usually more possibly preferred than those without them [9]. **Figure 14** demonstrates the photo with some semantic meanings. In **Figure 14(a)**, there are many human faces detected. Actually, people prefer to keep photos with faces, animals, and so on. Currently, some object detection methods, such as an AdaBoost algorithm, are able to detect photos with face, eyes, vehicles, and animals. However, most of the object detection methods work primarily on rigid objects. For nonrigid objects, in **Figure 14(b)**, both the photo segmentation and visual word techniques [4] are adopted to classify the regions of a photo as certain elements, say sky, water, tree, grass, roads, or buildings.
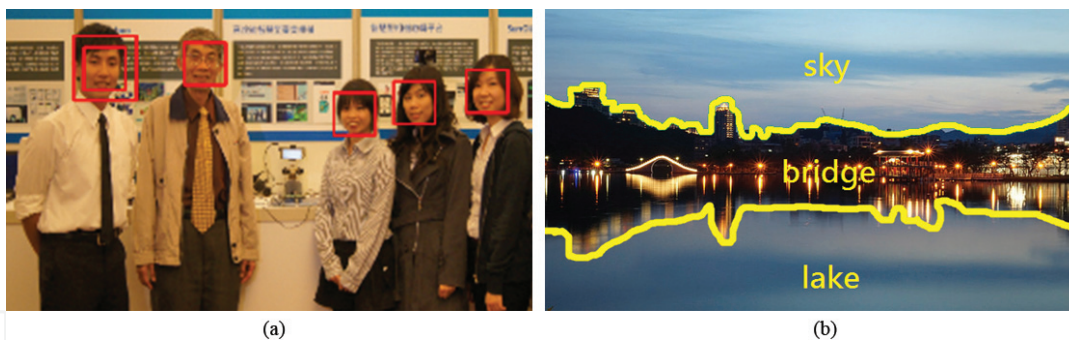

(a)   (b)

Figure 14. The semantics of a photo: (a) faces found by object detection methods; (b) the photo layout found by segmentation and visual word techniques.

5. Image composition

A visually pleasing photo usually has a good composition [5]. If a photo has certain composition characteristics, it is usually more popular than a photo without those. **Figure 15** shows some common types of image compositions, which are employed to perceive the beauty of a photo.

Salient regions and prominent lines are two important factors for analyzing the composition of a photo. The salient regions are the perceptually appealing areas, and the prominent lines are
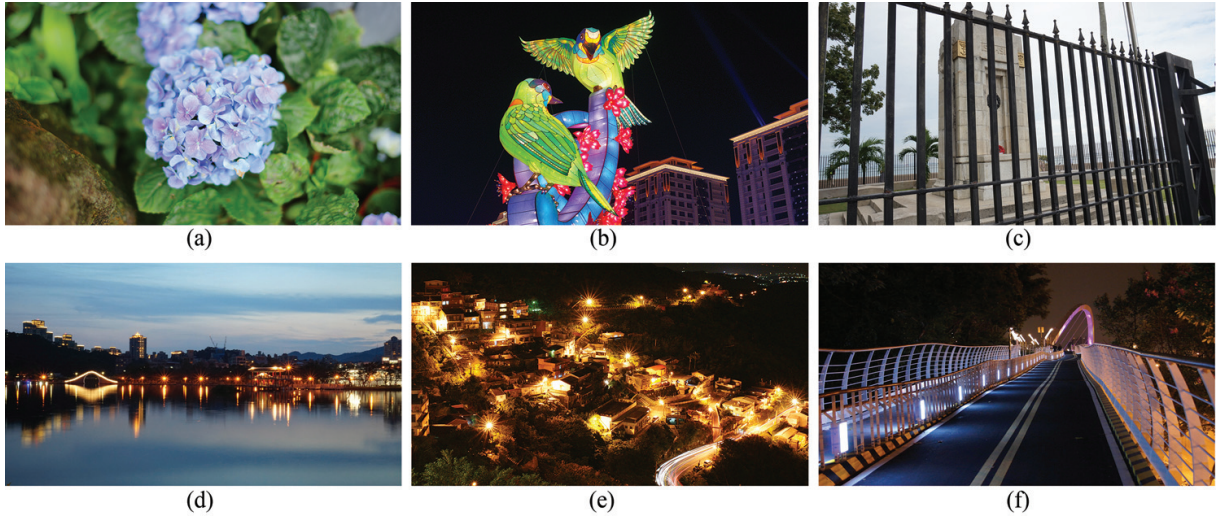
Figure 15. Six types of image compositions: (a) central; (b) rule of thirds; (c) vertical; (d) horizontal; (e) diagonal; (f) perspective.

visually existing edges. In this chapter, we take these two factors as the features fed to an artificial neural network for classifying the possible composition type of an input photo.

5.1. Salient map

Our attention is attracted to salient colors easily, which is a born ability of humans. This ability is important for complex biological systems to rapidly detect potential preys, predators, or mates in a visual world with cluttered objects. Therefore, by finding salient regions, it is possible to find a target object in a cluttered field of view. Locating the salient regions in a photo helps determine the composition of a photo. A salient map can be generated from calculating the salient degree of each color [7] as

$$S(I_k) = S(c_l) = \sum_{j=1}^{n} f_j D(c_l, c_j) \tag{13}$$

where I_k is the salient degree of pixel k, c_l is the color l in the CIELab color space, c_j is the color j in the CIELab color space, $D(x, y)$ is the color distance between two colors x and y in the CIELab color space, and f_j is the probability that color j appears in photo I. **Figure 16** illustrates an example of finding the salient map of a photo.

Figure 16. Illustration of finding the saliency map: (a) the original photo; (b) the grayscale photo of (a); (c) the saliency map of (a).
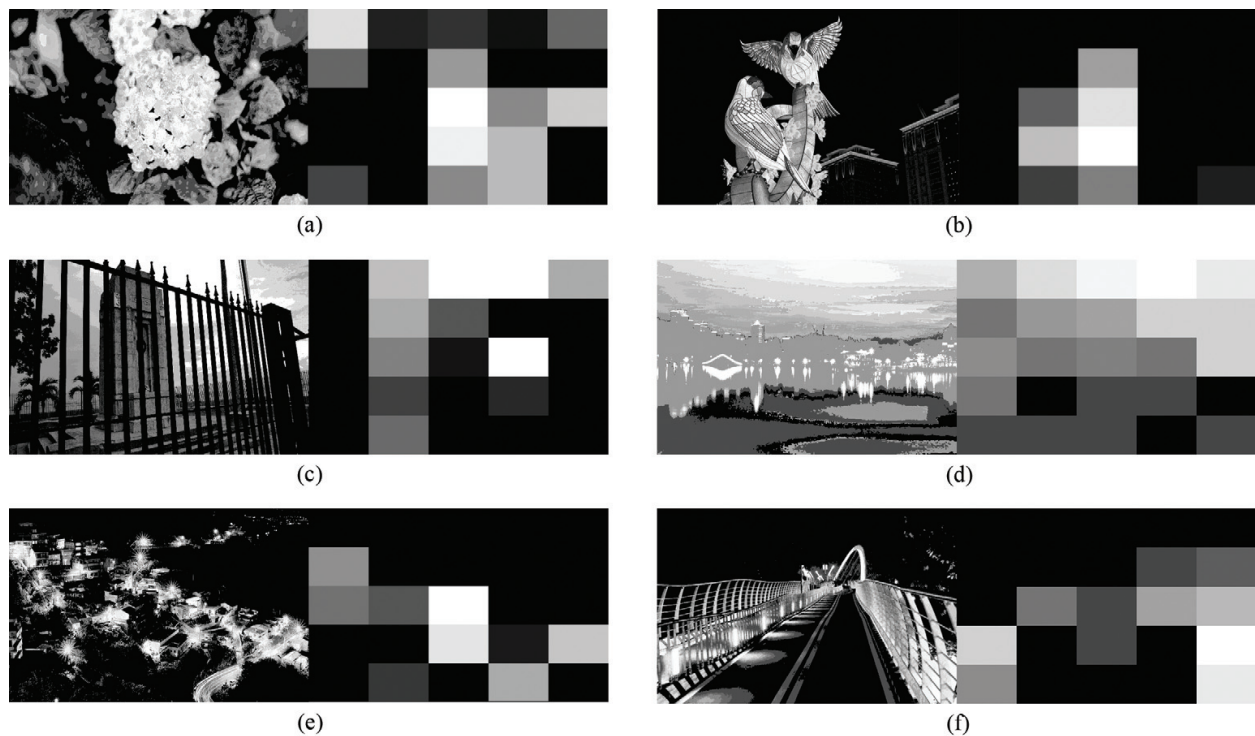
Figure 17. Illustration of a mosaic of 5x5 blocks resulting from the saliency map for each of six image compositions: (a) central; (b) rule of thirds; (c) vertical; (d) horizontal; (e) diagonal; (f) perspective.

The saliency map is further simplified into a mosaic of 5 x 5 blocks. The value of each block is calculated by averaging the salient degrees within the block. An illustration after simplification for each image composition is shown in **Figure 17**.

5.2. Prominent line

The Hough transform is used for finding prominent lines in a photo [8]. The prominent lines are the perceptual straight lines which appear in a photo. To detect the prominent lines, the edge detection must be performed first. The Canny edge detector is chosen in our proposed method. After the edge detection, prominent line detection is executed to detect straight lines in the photo, and the Hough transform is chosen as the detector. The concept of Hough transform is to transform the positions of all edge pixels in rectangular coordinates into polar coordinates, and select the transformed coordinates with more occurrences as the detected lines. What follows is the detailed procedure of prominent line detection.

Given a point $(p, q) = (rcos\theta, rsin\theta)$ on a line, let (x, y) be the other points on the line. Then

$$\frac{\Delta y}{\Delta x} = \frac{y - q}{x - p} = \frac{y - r\,sin\theta}{x - r\,cos\theta} \tag{14}$$

Because the slope of the line perpendicular to a straight line can be represented with $tan\theta$, the slope of the straight line is:

$$-\frac{1}{tan\theta} = -\frac{cos\theta}{sin\theta} \tag{15}$$

Combining the above two equations yields

$$\frac{y - r\,sin\theta}{x - r\,cos\theta} = -\frac{cos\theta}{sin\theta} \tag{16}$$

And the resulting equation can be rewritten as

$$y\,sin\theta - rsin^2\theta = -xcos\theta + rcos^2\theta \tag{17}$$

Through some mathematical manipulations, the equation of the straight line becomes

$$x\,cos\theta + y\,sin\theta = r \tag{18}$$

By substituting the coordinates x and y of every pixel located in the edges to the above equation, many possible combinations of r and θ are acquired, where the range of r is $0 < r \leq \sqrt{width^2 + height^2}$ and the range of θ is $-90° < \theta \leq 90°$. Therefore, we choose the combinations whose occurrences exceed a given threshold. The chosen combinations correspond to the prominent lines.

The results obtained from the prominent line detection performed on the photos of different compositions are demonstrated in **Figure 18** associated with the histograms of detected line orientations, respectively. In each histogram, the scope of 180 angle degrees is uniformly partitioned into 10 bins; that is, each bin contains 18 angle degrees. In the horizontal axis of a
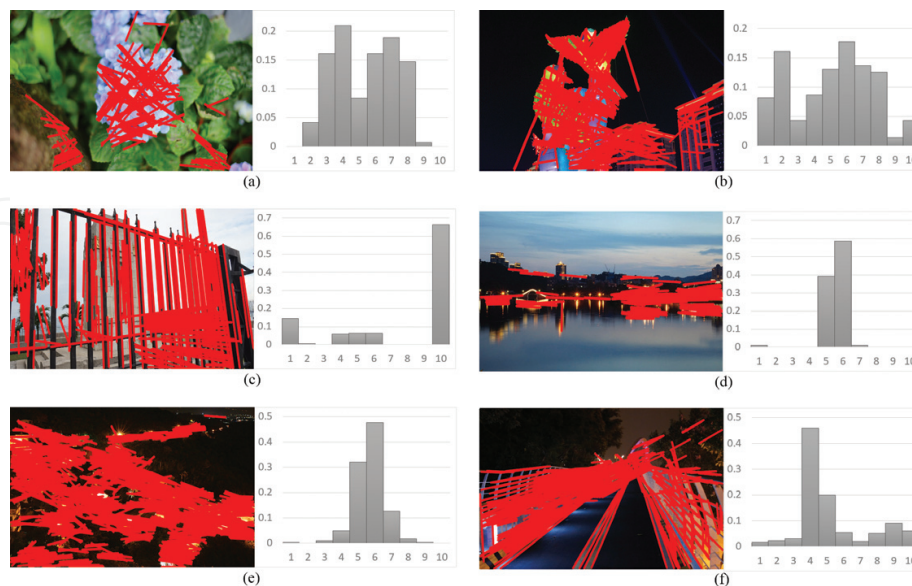

Figure 18. The detection results of the prominent lines appearing in six photos of different image compositions associated with their respective histograms of detected lines: (a) central; (b) rule of thirds; (c) vertical; (d) horizontal; (e) diagonal; (f) perspective.

histogram, bin 1 represents $-89°$ to $-72°$, bin 2 represents $-71°$ to $-54°$, …, and bin 10 represents $73°$ to $90°$. The vertical axis of the histogram means the percentages of the 10 orientations of the prominent lines appearing in a photo.

6. Machine learning

The machine learning method can be used to determine whether a photo is favorable or not. Basically, there are two kinds of machine learning methods grouped into supervised and unsupervised. In supervised machine learning, each sample photo has a label that indicates whether it is beautiful when training, while in unsupervised learning, there is no label. Choosing different machine learning methods leads to distinct functionalities and results.

6.1. Comparison of machine learning methods

In the unsupervised methods, the K-means algorithm is useful for clustering the n-dimensional feature vectors extracted from a photo. After performing the K-means algorithm and examining each cluster, there will be groups divided into three possible types that stand for favorable, unfavorable, and ambiguous classes. Then reserve the former two clusters. When a new photo is inputted, compute the distance between the centers of these two clusters, and an aesthetic score of the photo can be determined. On the other hand, two models can be used in supervised learning, including decision trees and neural networks. The benefit of a decision tree is that the classification rules are readable, which can be used to tell why a photo is favorable according to the machine learning result. An example of the decision tree is shown in **Figure 19**. The nodes in the decision tree can be either all low-level features as shown in **Figure 19(a)** or the nodes can comprise some semantics ones as shown in **Figure 19(b)**.

With the neural network approach, it is possible that the aesthetics score of a photo can be measured. There are two neurons in the output layer. The summation of the two output neurons' values is exactly one. The values of the "high quality" and "low quality" neurons
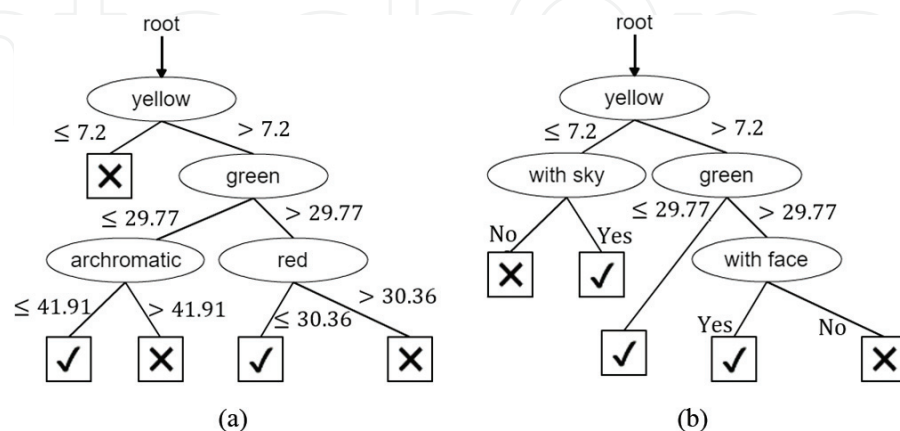

Figure 19. A simplified decision tree for the perception of beauty trained by 100 samples: (a) without semantics nodes; (b) with some semantics nodes.

are their respective probabilities. The structure of such a neural network for perceiving the beauty of a photo is illustrated in **Figure 20**.

6.2. Decision tree

A decision tree [10] is a popular supervised learning approach because the decision process is made from walking through a path of the tree and each path can be written as a readable classification rule. In a decision tree, the internal nodes excluding the leaf nodes represent features and their child edges are predicates for the features, such as "is larger than" and "is less than." A node without any children is a leaf node. The class labels are placed on the leaf nodes of the decision tree. When a series of feature values is fed to a decision tree, a path can be established through the decision tree from the root, via some internal nodes, and finally arrives at a leaf node. The label of the leaf node in the path is the classification result.

Decision tree algorithms are based on information theory, where the main idea is to calculate the entropy of the classes of data when the data are composed of specified features and branching values. The entropy of the entire data is computed by

$$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i),\tag{19}$$

where D is the input data, p_i is the percentage of class i that appears in all of the data and m is the number of classes. After using feature F to split D into v partitions, the information needed to classify D is computed by

$$Info_F(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)\tag{20}$$

where $|D|$ is the cardinality of data, $|D_j|$ is the cardinality of data for partition j, and $Info(D_j)$ comprises the entropy data for segment partition j.
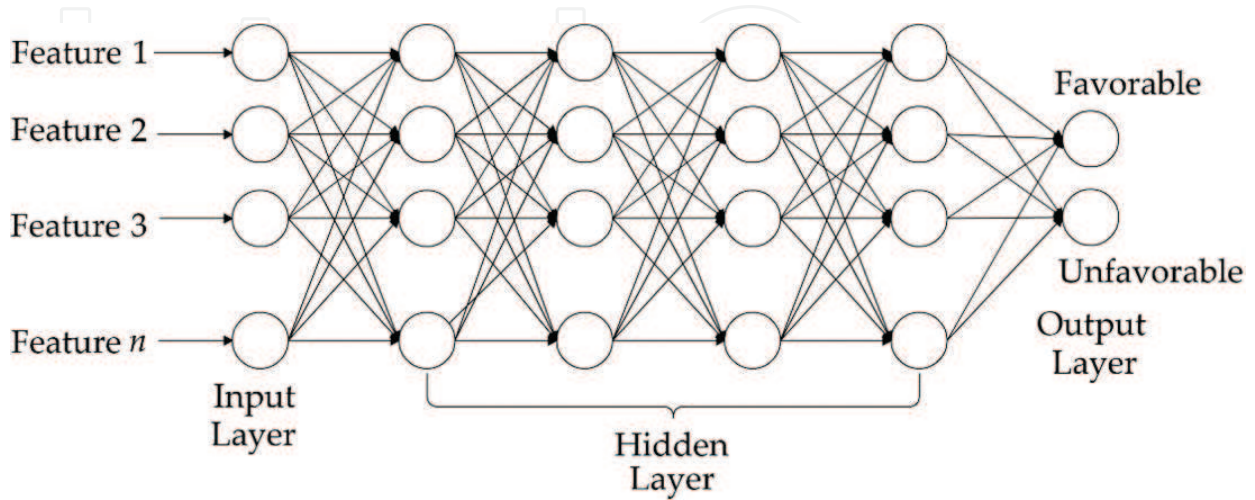

Figure 20. The structure of a neural network for perceiving the beauty of a photo.

The value of v is set to 2 if we want to build a binary decision tree in which the degree of each node is exactly two except leaf nodes. To split the input data into two partitions, a threshold is given initially. To find the optimal threshold, each distinct value in the data of the selected feature F is computed iteratively.

The information gain for feature F is defined as follows

$$\text{Gain}(F) = Info(D) - Info_F(D) \tag{21}$$

The feature F with the highest information gain is then selected as the feature for splitting the data. Nevertheless, the information gain tends to be biased toward the features with more levels. For example, if the brightness has 256 levels and the sharpness has 128 levels, then the result is often biased toward the brightness. Therefore, a gain ratio is used to normalize the information gain to eliminate bias, which is expressed below

$$\text{Gain}_{ratio(F)} = \frac{Gain(F)}{SplitInfo_F(D)}, \tag{22}$$

where

$$Split\ Info_F(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times log_2\left(\frac{|D_j|}{|D|}\right). \tag{23}$$

The $\text{Gain}_{ratio\ (F)}$ is adopted to replace the information gain to prevent bias toward the features with more levels.

A major problem that affects the performance of decision trees is over-fitting. If the depth of the tree is too high, some unnecessary nodes are then produced, which reduce the accuracy of the decision tree. As a result, pruning must be applied to the tree. A postpruning method involves trial and error. If a node is pruned by replacing it with a class leaf and the accuracy of the tree is better after the replacement, then the pruning is accepted; otherwise, keep the original sub-tree. An example is shown in **Figure 21**, where if the accuracy of **Figure 21(b)** is
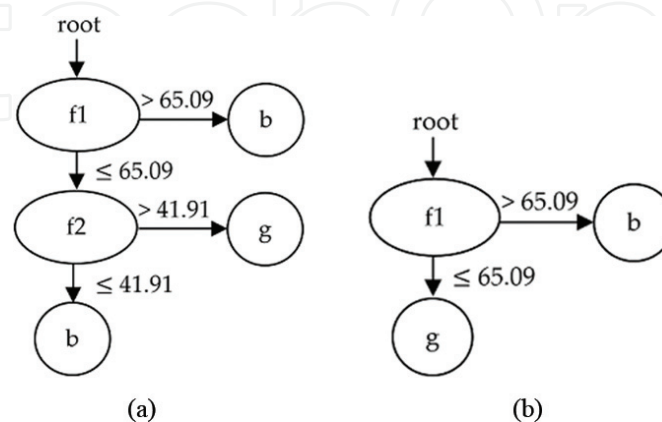

(a)   (b)

Figure 21. Pruning of a decision tree: (a) the original tree; (b) the tree has been pruned by replacing a sub-tree with a class leaf.

better than **Figure 21(a)**, then replace the original tree with the pruned tree in **Figure 21(b)**; otherwise, keep the unpruned tree in **Figure 21(a)**.

6.3. Artificial neural network

The multilayer perceptron (MLP) is a feed-forward neural network whose architecture is composed of three main substructures, namely the input, hidden, and output layers. **Figure 22** shows the fundamental architecture of an MLP.

The MLP comprises various neurons and synapses associated with connection weights, in which the output of a neuron is derived from an activation function via the weighted sum of its inputs. During the training, the weights are randomly initialized within a range. When the output of a neuron is different from an expected value, the weights are adjusted iteratively until their quantities are almost unchanged.

A neuron in the hidden and output layers is activated by applying inputs $x_1(p)$, $x_2(p)$, ... ,$x_n(p)$ at iteration p, which are weighted by $w_1(p)$, $w_2(p)$, ... ,$w_n(p)$, respectively. For instance, the sigmoid function serves as the activation function of the neuron, which is expressed as follows

$$y(x) = Y^{sigmoid}\left[\sum_{i=1}^{n} x_i(p) \cdot w_i(p) - \theta\right] = 1/1 + \exp\left(-\sum_{i=1}^{n} x_i(p) \cdot w_i(p) - \theta\right) \quad (24)$$

where θ is a given bias acting as the quantity deviated from the original input of the neuron.

However, the weighted sum can only solve linear problems. To overcome linear inseparability, a hidden layer is added to constitute the MLP. Because such a neural network is trained with supervised learning, a back-propagation algorithm is developed for updating the weights from the output layer to the input layer [11]. Once all the weights have been trained, the MLP can be employed to predict the output immediately when an input is fed.

Figure 23 shows a three-layer perceptron comprising a hidden layer, which requires the back-propagation mechanism (algorithm) to update the weights in the course of training.
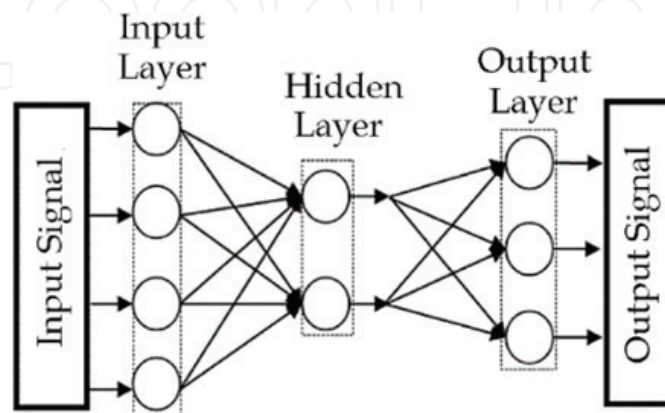

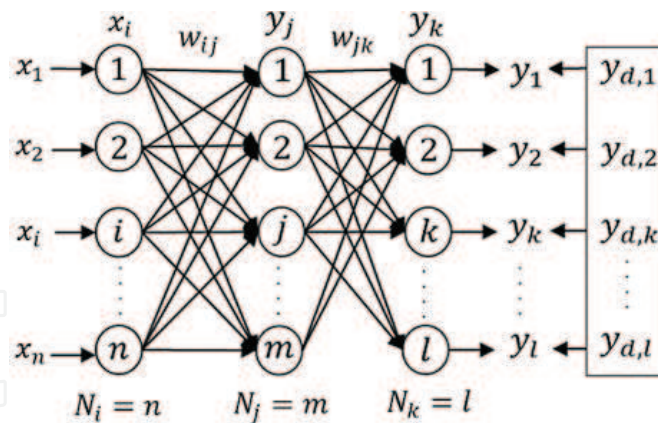**Figure 22.** Example of the architecture of a multilayer perceptron.

Figure 23. Three-layered back-propagation neural network.

In **Figure 23**, N_i, N_j, and N_k are the numbers of neurons in the input, hidden, and output layers of the network, respectively. The weights w_{ij} and w_{jk} and biases are initialized by taking the random numbers that are uniformly distributed within a small empirical range, say $\left(-\frac{2.4}{N_i}, \frac{2.4}{N_i}\right)$.

Both the weights w_{ij} and w_{jk} are further updated through the delta updating rule depicted below. At iteration p, w_{ij}, and w_{jk} are adjusted by Δw_{ij} (p) and Δw_{jk} (p) according to the following formulas

$$\Delta w_{ij}(p) = \alpha x_i(p)\delta_j(p) \text{ and } \Delta w_{jk}(p) = \alpha x_j(p)\delta_k(p) \tag{25}$$

where α is the learning rate ranged from 0.1 to 0.5, and δ_j (p) and δ_k (p) are the error gradients for the hidden and output layers, respectively.

Subsequently, the weights w_{ij} and w_{jk} at iteration p + 1 are calculated as follows:

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \text{ and } w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \tag{26}$$

Following are the respective inputs y_j and y_k for neuron j and neuron k in the hidden and output layers at iteration p:

$$y_j(p) = Y^{sigmoid}\left[\sum_{i=1}^{n} x_i(p)w_{ij}(p) - \theta_j\right], j = 1, 2, ..., m \tag{27}$$

$$y_k(p) = Y^{sigmoid}\left[\sum_{j=1}^{m} y_j(p)w_{jk}(p) - \theta_k\right], k = 1, 2, ..., l \tag{28}$$

Where θ_j and θ_k stand for the input biases of neurons **j** and **k**, respectively.

The error between the desired value and the value predicted by the three-layer perceptron is obtained from

$$e_k(p) = y_{d,k}(p) - y_k(p) \tag{29}$$

where $y_{d,k}(p)$ and $y_k(p)$ are the desired and predicted outputs, respectively.

The error gradient for the output layer is computed as follows

$$\delta_k(p) = y_k(p)[1 - y_k(p)]e_k(p) \tag{30}$$

The weights between the hidden layer and the output layer are adjusted by

$$\Delta w_{jk}(p) = \alpha y_j(p)\delta_k(p) = \alpha y_j(p)e_k(p)y_k(p)[1 - y_k(p)] \tag{31}$$

Thus, the error for the output of the output layer can be propagated back to the hidden layer, and the error for the output of the hidden layer is computed as follows

$$e_j(p) = \sum_{k=1}^{1} \delta_k(p)w_{jk}(p) \tag{32}$$

The error gradient for the hidden layer is formulated below

$$\delta_j(p) = y_j(p)[1 - y_j(p)]e_j(p) \tag{33}$$

The weights between the input layer and the hidden layer are computed by

$$\Delta w_{ij}(p) = \alpha y_i(p)\delta_j(p) = \alpha y_i(p)e_j(p)y_j(p)[1 - y_j(p)] \tag{34}$$

Thus, the iteration p is increased by 1, and the procedure is repeated until the sum of squared errors is sufficiently small or the number of iterations reaches a given maximum. The following defines the sum of squared errors

$$E = \frac{1}{2}\sum_{p=1}^{N_T}\sum_{k=1}^{l}(y_{d,k}(p) - y_k(p))^2 \tag{35}$$

where N_T is the number of training samples and l is the number of neurons in the output layer. In our proposed method, such a three-layer perceptron is applied to classifying the type of image composition for an input photo.

7. Personal preference

Photo aesthetics is subjective to different groups of people. To deal with this problem, we adopt social networks to collect people's preferences; for instance, the attributes of personal information and the features of his/her favorite pictures. The correlation between the attributes of people and photo features is calculated. A bias is used to influence one of the feature values obtained from different people, which is formulated as follows

$$\hat{b}_i = b_i + \sum_{j=1}^{n} norm(p_j)\gamma_{ij}\,\mu_i \tag{36}$$

where I is the index for a photo feature (brightness, color contrast, etc.), j is the index for the personal attribute (gender, age, education, etc.), and n is the number of personal attributes. Besides, b_i is the original bias for each feature value, p_j is a photographer's attribute value, $norm(p_j)$ is a normalized attribute value (from 0 to 1), γ_{ij} is the correlation for a photo feature value and a photographer's attribute value, which is ranged from -1 to $+1$, and μ_i is the personal influence for a feature value. The bias value can be used for the decision tree described in the previous section.

8. Experimental results

In this chapter, the performance of two computational aesthetics manners for the perception of beauty is evaluated, which are based on image composition analysis and low-level features to determine whether a photo meets the criterion of a professional photographing via different classifiers. The parameters for the classifiers are depicted as follows: for the support vector machine (SVM), the radial basis function (RBF) is chosen as the kernel function, and the cost is set to 1. For the MLP, the number of neurons in the hidden layer is set to half the sum of the numbers of features and classes, which is equal to 8 for the first experiment and 22 for the second experiment. The learning rate is 0.3, and the number of iterations is 500 during the training. For the Radial basis network, the minimum standard deviation is set to 0.1, clustering seed is 1, the number of clusters is 2, and the ridge is 10^{-8}. For the AdaBoost algorithm, the weak classifiers are decision stump and the number of iterations is set to 10, the seed is set to 1, and the weight of threshold is set to 100. For the decision tree J48, the confidence factor is set to 0.25. For the random forest, the number of trees is set to 100.

8.1. Test on low-level features used for perceiving the beauty of photos

In this experiment, we choose multiple low-level features to classify whether a photo is favorable or not automatically. In total, 15,000 photos are collected, and 13 features are extracted from them, which include color components (red, green, blue, cyan, magenta, and yellow), sharpness, brightness, contrast, saturation, color balance, colorfulness, and simplicity. Each of the photos is marked as favorable or unfavorable for training. The testing is performed under 10-fold verification.

Many actual photos are examined in the photo beauty measurement system, including true positive samples (favorable photos classified as favorable; correct result), false negative samples (favorable photos classified as unfavorable; incorrect result), false positive samples (unfavorable photos classified as favorable; incorrect result), true negative samples (unfavorable photos classified as unfavorable; correct result). The classification results of some sample photos are as shown in **Figure 24** where **Figures 24(a)** and **24(d)** are correct results. **Figure 24(b)** shows the photos whose features should be salient enough but they are determined as amateur
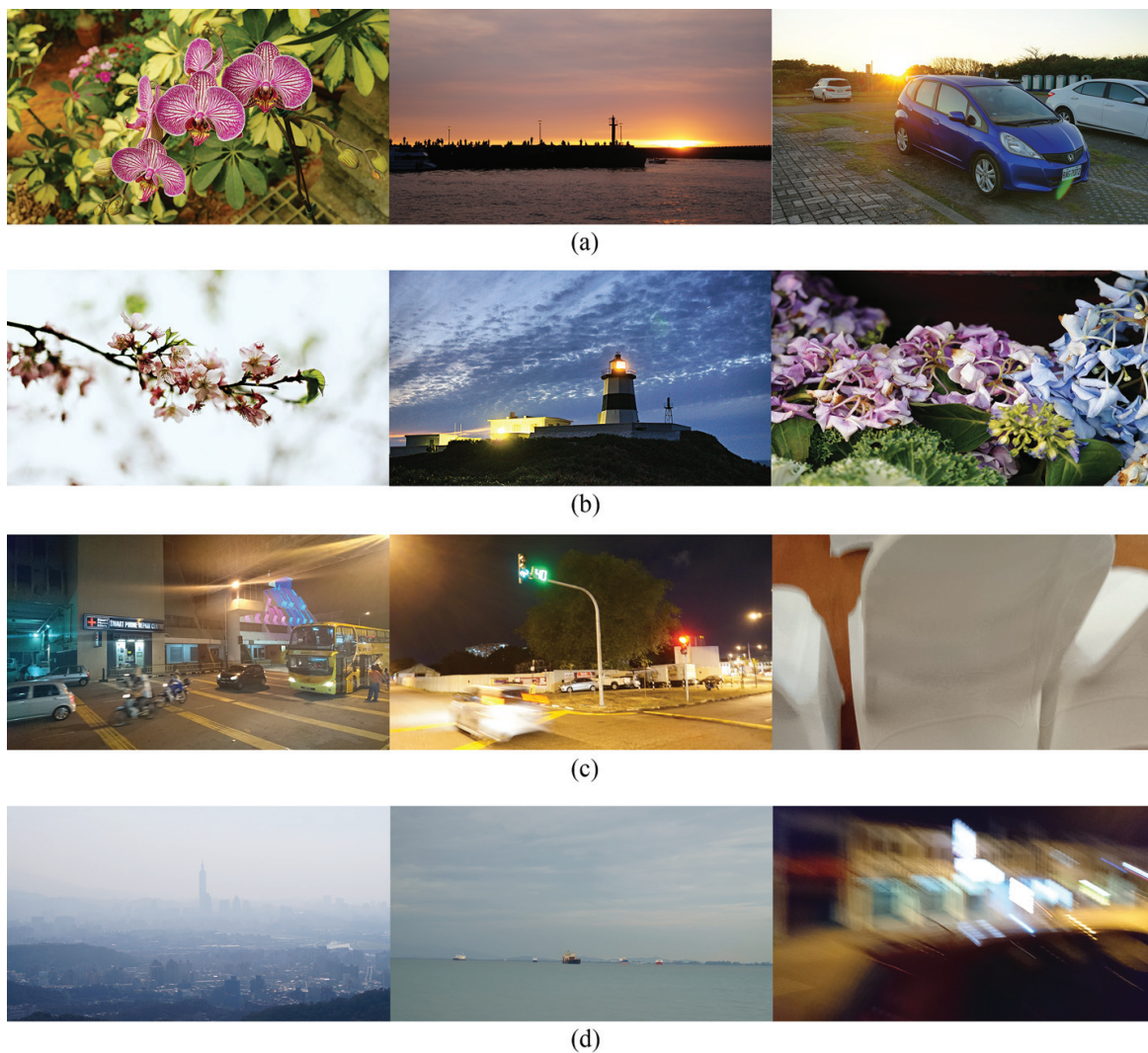
(a)

(b)

(c)

(d)

Figure 24. Four classification results of some sample photos: (a) true positive; (b) false negative; (c) false positive; (d) true negative.

and unfavorable. **Figure 24(c)** shows that human knowledge about photo contents should also be applied; however, they are still recognized as favorable in spite of ill problems appearing in the contents.

Table 1 shows both the accuracy and the area under a ROC curve (AUC) of classifying whether a photo meets the condition of a beautiful photo. Compared to other classifiers, the MLP and tree-based ones have better performance. The MLP can be used to show the aesthetic score of a photo, while the decision tree J48 is able to generate readable rules of the classifier.

8.2. Test on image composition analysis for perceiving the beauty of photos

In this experiment, the image composition analysis is tested. Thirty five features are used for training an MLP, including a stack vector of 25 salient region values depicted in Section 5.1, and the angle degrees of prominent lines are ranged from $-90°$ to $90°$ where every 18 angle degrees results in a bin. Therefore, the numbers of prominent lines are counted in 10 bins of the

Classifier	Accuracy (%)	AUC (%)	Classifier	Accuracy (%)	AUC (%)
AdaBoost algorithm	82.3	73.4	Multilayer perceptron	91.5	95.8
Radial basis network	78.0	83.5	Decision tree J48	94.6	94.1
Support vector machine	81.0	73.4	Random forest	95.8	98.9

Table 1. The AUC and accuracy of different classifiers.

angle degrees, which act as the remaining features described in Section 5.2. In consequence, some photo samples are provided to illustrate the performance of image composition analysis. **Figure 25** shows the correctly classified samples whereas **Figure 26** shows the incorrectly classified ones.
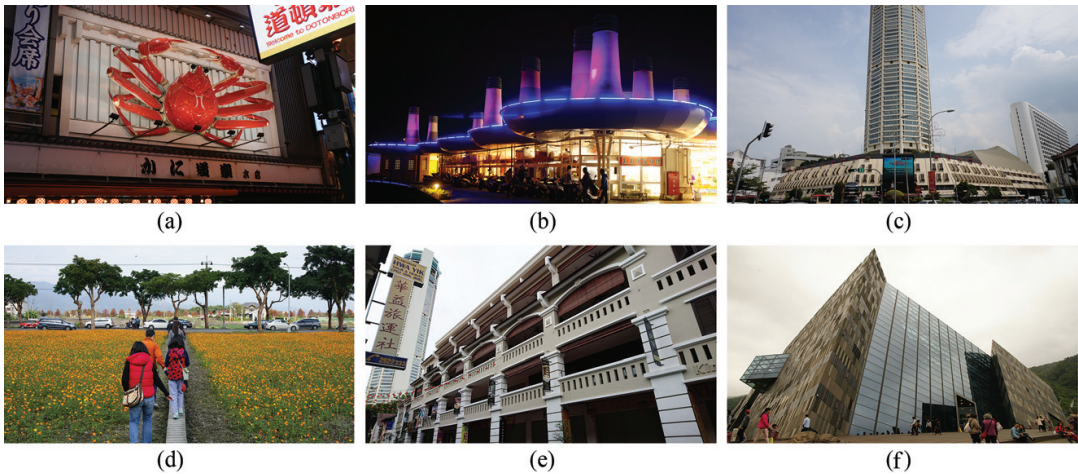

Figure 25. Correctly classified image compositions: (a) central; (b) rule of thirds; (c) vertical; (d) horizontal; (e) diagonal; (f) perspective.
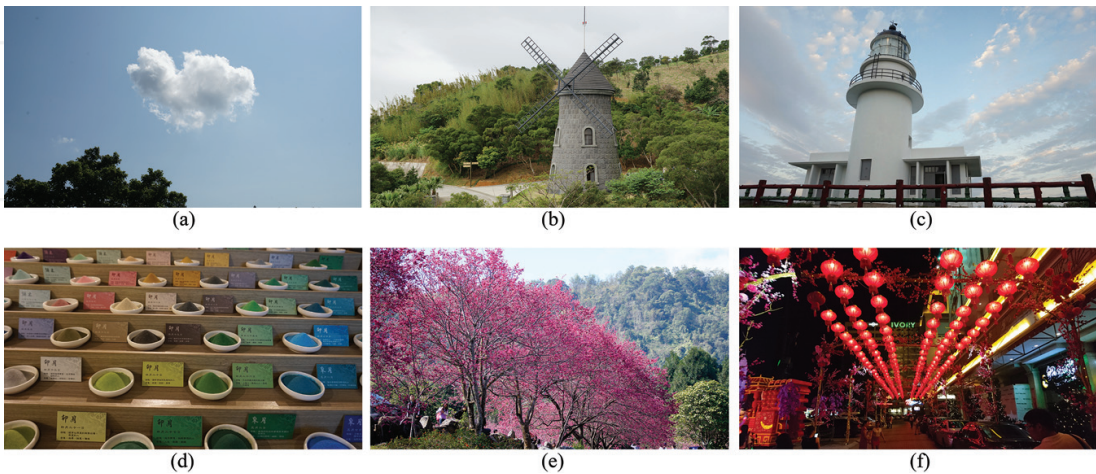

Figure 26. Incorrectly classified image compositions: (a) central misclassified as horizontal; (b) rule of thirds misclassified as horizontal; (c) vertical misclassified as horizontal; (d) horizontal misclassified as rule of thirds; (e) diagonal misclassified as horizontal; (f) perspective misclassified as rule of thirds.

Classifier	Accuracy (%)	Classifier	Accuracy (%)
AdaBoost algorithm	68.7	Multilayer perceptron	96.7
Radial basis network	79.1	Decision tree J48	94.2
Support vector machine	61.7	Random forest	97.2

Table 2. The accuracy of image composition analysis using different classifiers.

Composition	AUC (%)	Composition	AUC (%)
Center	94.5	Rule of thirds	83.9
Diagonal	93.9	Perspective	90.1
Horizontal	92.5	Vertical	98.9

Table 3. The AUC of multiple image compositions.

From the incorrectly classified composition samples, we can see that horizontal and rule of thirds compositions are two commonly mistaken ones, which may be caused by horizontal lines existing in most photos and distractors often existing in one-third of photos. A solution is that the weights of horizontal and rule of thirds compositions can be lowered in the output layer of the MLP.

In **Table 2**, the accuracy of image composition classification is calculated by the percentage of correctly classified instances to all samples. Each of the classifiers is tested using 10-fold verification. Tree-based and MLP classifiers have higher performance than others.

Table 3 lists AUC of six image composition. The AUC is calculated by the percentage of the area under a ROC curve. The rule of thirds composition has the least AUC, because it is often confused with the center composition, while the perspective composition also has lower AUC, because it is frequently confused with the diagonal composition.

9. Conclusion

In this chapter, a measurement method of photo aesthetics is presented. Several factors for measuring the beauty of a photo are discussed, including low-level features, photo semantics, image composition, and personal preference. Image composition plays an important role on the photo beauty measurement and a detection method is presented in this chapter. Object detection and image segmentation algorithms are aided to illustrate the layout of a photo and the social network helps finding the personal preference of a photo. Both the decision tree and MLP have high accuracy that is above 90% for evaluation. The decision tree can generate readable classification rules, while the MLP can give aesthetics scores to stand for the degree of beauty. The photo beauty measurement system can be implemented in real time, which is suitable for the installation of various kinds of equipment.

Acknowledgements

Author details

Chin-Shyurng Fahn* and Meng-Luen Wu

*Address all correspondence to: csfahn@mail.ntust.edu.tw

Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei City, Taiwan

References

[1] Ke Y, Tang X, Jing F. The design of high-level features for photo quality assessment. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06); 17-22 June 2006; New York, NY, New York: IEEE; 2006. pp. 419-426

[2] Yeh CH, Ho YC, Barsky BA, Ouhyoung M. Personalized photograph ranking and selection system. In: Proceedings of the 18th ACM International Conference on Multimedia (MM '10); 25-29 October 2010; Firenze, Italy. New York: ACM; 2010. pp. 211-220

[3] Lo KY, Liu KH, Chen CS. Intelligent photographing interface with on-device aesthetic quality assessment. In: Proceedings of Asian Conference on Computer Vision (ACCV '12); 5-9 November 2012; Daejeon, Korea. Berlin, Heidelberg: Springer; 2012. pp. 533-544

[4] Composition–Definition: Photographic Aesthetics–Photokonnexion [Internet]. 2010. Available from: http://www.photokonnexion.com/3269-2/ [Accessed: November 09, 2016]

[5] Wu ML, Fahn CS. A decision tree based image enhancement instruction system for producing contemporary style images. In: Proceedings of International Conference on Human-Computer Interaction (HCII '16); 19-22 July 2016; Toronto, Canada. Cham, Switzerland: Springer; 2016. pp. 80-90

[6] Bhattacharya S, Sukthankar R, Shah M. A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the 18th ACM International Conference on Multimedia (MM '10); 25-29 October 2010; Firenze, Italy. New York: ACM; 2010. pp. 271-280

[7] Chen MM, Mitra NJ, Huang X, Torr PHS, Hu SM. Global contrast based salient region detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015;**37**(3): 569-582. DOI: 10.1109/TPAMI.2014.2345401

[8]   Duda RO, Har PE. Use of the Hough transformation to detect lines and curves in picture. Communications of the ACM. 1971;**15**(1):11-15. DOI: 10.1145/361237.36124

[9]   Yang J, Jiang YG, Hauptmann AG, Ngo CW. Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the International Workshop on Multimedia Information Retrieval (MM '07); 24-29 September 2007; Bavaria, Germany. New York, NY: ACM; 2007. pp. 197-206

[10] Apté C, Weiss S. Data mining with decision trees and decision rule. Future Generation Computer System. 1997;**13**(2-3):197-210. DOI: 10.1002/9781118029145.ch6

[11] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;**323**(6088):533-536. DOI:10.1038/323533a0