Computational Aesthetic Measurement of Photographs Based on Multi-features with Saliency

Yimin Zhou[1], Yunlan Tan[1,2], and Guangyao Li[1]

[1] College of Electronics and Information
Tongji University
Shanghai, 201804, China
[2] School of Electronics and Information
Jinggangshan University
Ji'an, Jiangxi, 201804, China
{1010080053,2011tan,lgy}@tongji.edu.cn

Abstract. Based on the existent computational aesthetic measurements, we present a new approach that combining both saliency region detection and extraction with a feature set in line with the principle of human vision. We first extract the saliency region using frequency-based method, then extract 53 features from both local and global regions, and select top 15 features which can determine the best aesthetic value. We run both SVM classification & regression and CART as well as linear regression on the filtered dataset. The experiments show a meaningful result of an accuracy above 70%.

Keywords: saliency features, computational aesthetic measurement, aesthetics of photography, SVM classification and regression, CART.

1 Introduction

With the emergence of computer, internet and digital camera, photographs acquisition, storage and sharing become even more convenient. Aesthetics of photography is how people assess the beauty of one photograph. A professional photographer may decompose it into parts, then judge each in detail, even evaluate brightness, color combination and contrast etc., which are integrated factors in addition to some standards of the photography industry. All these kinds of assessment are much largely dominated by the particular individual who engaged in the job. According to Flickr's statistics[1], users upload about 6.5 million photographs every day. It is a heavy challenge to make aesthetics measurement of such huge amount of photographs. Nevertheless, computer, again, plays the central role in dealing with the situation. To tackle this knotty problem, in this paper, we present a computational and experimental approach to handle it in a statistical learning way. This allows us to exclude those irrelevant factors of photographs and focus on only those certain key features that can identify ones of high aesthetics from the vast amount of the collection of albums on

D.-S. Huang et al. (Eds.): ICIC 2014, LNCS 8588, pp. 357–366, 2014.

the internet in a statistical sense. In the age of information explosion as well as the digital image information, to develop intelligent systems for automatic evaluation of digital photographs is much critically demanded.

In our work, we present a process of images' related features detection, calculation, training, classification and prediction which is a variant of Datta's[2]. Unlike Datta's, we consider both image's saliency region and traditional aesthetic feature including Kolmogorov complexity, improved wavelet feature and NSCT wavelet decomposition etc. We achieve our experiment results with promising aesthetic measurement' values based on authoritative online photograph galleries. The accuracy close to those of Datta's[2] and other former researchers proves our method's availability and validity. The remainder of this article is organized as follows. In the next section, we review related work; In section 3, we illustrate the method of image feature extraction, training and classification; In section 4, we present the experimental results and some insight from them; In section 5, a conclusion of our work has been drawn.

2 Related Works

Measurement of the aesthetics value can be retraced to the writing of American Mathematician Birkhoff in 1933, the "Aesthetic Measure"[3]. In this work, he presented the formula of aesthetics measurement,

$$M = \frac{O}{C} \tag{1}$$

Where M means measurement, O represents order and C denotes complexity. Machado et al.[4] treated aesthetics measurement as proportional image complexity and inversely proportional to processing complexity of the image in brain. They gave the formula of aesthetic measurement of the image,

$$Measure = \frac{IC}{PC} \tag{2}$$

IC is the ratio of JPEG compressed image's error and the compression ratio, while PC is fractal image compressibility. Rigau et al.[5-6] combined information theory, which extended Birkhoff's theory by applying colors' distribution, Shannon's Entropy and Kolmogorov complexity. According to their framework, Juan Romero and Penousal Machado[7] explored the use of image compression estimates to predict the aesthetic merit of the images.

Although a theoretical measure of computing the aesthetics value is a fast and accurate way, classical assessments of features such as similarity, balance, combination, orientation, hue, harmony and contrast, still dominate most observers and connoisseurs' sensory. Wang and Datta et al.[2]from Pennsylvania were the first to realize the quantization of aesthetics measurement of image features. They extracted image features from the images, including the brightness, color distribution, wavelet, region composition and depth of field, and then applied SVM or linear regression to classify the high from the low quality photographs. They achieved an accuracy of 70.12%. The ACQUINE[8] aesthetics value measurement system developed by them is a typi-

cal aesthetics evaluation and search engine. Wu et al.[9] extended the SVM classification method utilized by Datta and Wang to predict aesthetics measure values. Ke et al.[10] distinguished professional high quality images from low quality snapshots by evaluating the aesthetics value from images' visual features. But, they ignored the differentiation between the visual attention regions and the remains. Wong et al.[11] presented saliency-enhanced image classification method but with a partial-automatic saliency detection approach and a relatively stale feature sets. Luo et al.[12] exploited the blur detection to estimate roughly focused on main area's features. Subhabrata et al.[13] discussed the rules of thirds and visual weight balance in detail and developed an interactive application that enabled users to improve the visual aesthetics of their digital photographs using spatial re-composition.

3 Aesthetic Classification and Sorting

3.1 Frequency-Tuned Saliency Region Detection

Frequency-tuned saliency-region detection was presented by Achanta[14]. Figure 1 illustrates one of the source images and its detected saliency region.Every RGB image is converted to HSV color space, generating the two-dimensional matrices I_H, I_S, I_V, each size of $X \times Y$. In comparison to our saliency method with the former largest patches selection method, we extract both features of saliency and largest patches to demonstrate that saliency features are dominant than the largest patches features.

Fig. 1. (a)Source image (ID. 65200) (b) the corresponding result of saliency detection (c) The related top 5 patches detected

Altogether we extract 53 features of local and global features for every selected image. The first selected features are referred as candidate ones for further refinement and are denoted as $F=\{ f_i \mid 1 \le i \le 53 \}$ that are described as follows.

3.2 Global Features Extraction

- **Aspect Ratio.** As is known to all, the value approximating 4:3 and 16:9, which are the golden ratio and chosen as television screens, are much related to the viewing pleasure. The aspect ratio features is

$$f_1 = \text{Width/Height}$$

Brightness. We use the average pixel intensity as the brightness of the image, which is defined as f_2. The saliency brightness is defined as f_3, which represents the average pixel intensity of the saliency region.

$$f_2=\frac{1}{XY}\sum_{x=o}^{X-1}\sum_{y=0}^{Y-1}I_V(x,y),\quad f_3=\frac{1}{XY}\sum_{(x,y)\in saliency}I_V(x,y)$$

- **Saturation.** Saturation or as its definition, subjective experience of vividness or richness of color indicates chromatic purity. We compute the average saturation as

$$f_4=\frac{1}{XY}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}I_s(x,y)$$

$$f_5=\frac{1}{XY}\sum_{(x,y)\in saliency}I_S(x,y)$$

- **Dark Channel .** Dark channel was introduced by He et al.[15] for haze removal. The value of dark channel[16] of an image I is defined as

$$I_{\text{dark}}(\text{i})=\min_{c\in R,G,B}(\min_{i'\in\Omega(i)}I_c(i'))$$

Where I_c is the color channel of I and $\Omega(i)$ is the neighborhood of pixel i. The $\Omega(i)$ is chosen as 11×11 local patch. The dark channel feature of the photograph I is computed as the average of the normalized dark channel values in the area of the whole image

$$f_6=\frac{1}{\|S\|}\sum_{(i)\in S}\frac{I_{dark}(i)}{\sum_{c\in R,G,B}I_C(i)}$$

Where S is the subject area of the photograph.

- **Kolmogorov Complexity**. Kolmogorov[17] [5] complexity is defined as the normalization of the order

$$f_7=\text{M}_\text{k}=\frac{NH_{\max}-K}{NH_{\max}}$$

M_k takes the values within [0,1] and expresses the photograph's degree of order without any prior knowledge of its size. Here NH_{max} is the original size of the photograph and K is the Kolmogorov complexity of the compressed size. JPEG compressor is selected because of its ability to discover patterns while losing little information

which is imperceptible to the human eyes. We also compute the Kolmogorov complexity of the saliency region and the corresponding feature f_8 in a similar way.

- **NSCT Texture Features.** We apply a non-subsampled contourlet transform (NSCT)[18] to the HSV color space of the image and use these samples as our selected features. The Non-subsampled Directional Filter Bank(NSDFB) is constructed by combining critically-sampled two-channel fan filter banks and resample operations. The result is a tree-structured filter bank that splits the 2-D frequency plane into directional wedges. Figure 2 illustrates a three-level decomposition.

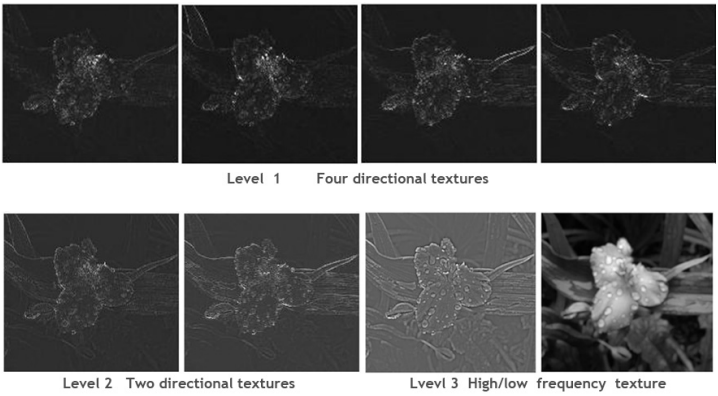

Fig. 2. The different level textures of NSCT on ID. 217257

We take the average NSCT wavelet transform of level 1 to 3 of HSV color space of each image as our texture features.

$$f_{8+i}=\frac{1}{XY}\sum_{W_H,S=1}(x,y)$$

$$f_{9+i}=\frac{1}{XY}\sum_{W_H,S=2,D_1}(x,y)+\sum_{W_H,S=2,D_2}(x,y)$$

$$f_{10+i}=\frac{1}{XY}\sum_{W_H,S=2,D_1}(x,y)+\sum_{W_H,S=2,D_2}(x,y)+\sum_{W_H,S=2,D_3}(x,y)$$

where i=1,11,15. So are S and V. We also define similar features for saliency region.

3.3 Local Features Extraction

- **Rule of Thirds.** Rule of Thirds is a very popular rule of thumb in photography. We consider the rule of thirds inside saliency region, so we do not omit the rule of thirds features. We compute the average hue of the inner thirds region as,

$$f_{27}=\frac{9}{XY}\sum_{x=X/3}^{2X/3}\sum_{y=Y/3}^{2Y/3}I_H(x,y)$$

f_{28}, f_{29} are computed similarly for I_s and I_v respectively.

- **The top5-Patches Features.** We still present the old top 5 patches related features for comparison with the saliency related features. The image is transformed in the LUV space because the perceived color changes well in this locally Euclidean distance model space. With a fixed threshold for all the photographs, we use a K-means algorithm after a K-center algorithm computes clusters. Following a connected component analysis, color-based segments are obtained. The largest segments formed are denoted as {s1,…,s5}. One sample and its top 5 patches are shown in Figure 1(c). Then we calculate the average hue, saturation and value for each of the top 5 patches as features 30-32, 34-36, 38-40, 42-44, 46-48. That is,

$$f_{i+30}=\frac{|S_i|}{(XY)}$$

i=1,5,9,13,17. And finally, the rough positions of each segment are stored as features f_{33}, f_{37}, f_{41}, f_{44}, f_{49}. The image is divided into 3 equal parts along horizontal and vertical directions, locate the centroid of each patch s_i in the block. f_{32+i}=(10r+c) where (r,c) $\in$ {(1,1),…,(3,3)} indicates the corresponding block starting with left and top.

- **Visual Attention Center.** The stress points[13] are the four points in an image which are the intersections of horizontal and vertical 1/3 and 2/3 dividing lines. We include four features of visual attention centers, which are represented by f_{50} through f_{53}.

$$f_{50}=\frac{1}{XY}\| v-s_1 \|_2^2$$

3.4 Training and Testing

SVM Classification and regression**.** All 53 features in $\mathcal{F}$ were extracted and normalized to the range of [0, 1] to form the data of experiment. We make a filter of aesthetics scores where greater than or equal to 5.8 being treated as high values and less than or equal to 4.2 being treated as low values. The scores between 4.2 and 5.8 are considered as ambiguous. People often don't fully distinguish high from low or vice versa. We then replicate the data to generate equal number of samples of the high and the low to ensure equal priors. We use the standard RBF kernel with the parameter γ=18.5 and cost=1.0 to perform our classification job. We chose livSVM as our algorithm package. SVM is run 10 times per feature by m-fold cross-validation. Then top 27 features were filtered out. We then use the greedy algorithm to stop after found the top 15 features and used them to build the SVM-based classifier. After doing this, we

try to predict the aesthetic values by using the SVM based linear regression by setting γ=200.0, ε=0.1 and cost=5.0 to have a moderate result.

CART Classification and regression**.** Unlike SVM, the benefit of CART is we construct a regression tree to predict aesthetic values in a fast and effective way. We chose the recursive partitioning implementation(RPART) as our CART building utility, and built a two-class classification tree model for the same training set in SVM turn.

Linear Regression. We perform linear regression in polynomial terms of the feature values to see whether it can directly predict the aesthetics scores from the feature vector or not. All data were used. For each feature f_i, the polynomial terms f_i^2, f_i^3, $f_i^{1/3}$, $f_i^{2/3}$ are used as independent variables.

4 Experimental Results and Analysis

We test the proposed filter algorithm on hundreds of color images depicting typical scenes of natural landscapes by a PC with Duo CPU 2.8GHz. Our experiments are programmed in c++ and MATlab.

4.1 Dataset

We use dpchallenge.com as our dataset source. The dataset contains user scores of 304,073 images in a grade from 1 to 10. In the collection, each photo has been evaluated by at least one hundred users. The collection contains several evaluation indices for each image, including the number of aesthetics ratings received, the means of ratings and a distribution of quality ratings on a 1-10 scale. We download 28,896 photographs from a miscellaneous contents set from its portal and chose 5,967 from them containing all ranges of aesthetic values every of which were voted by no less than two hundred viewers. Figure 3 shows the distribution of samples according to their aesthetic value.

From our experiences, only about one-fifth to one-fourth results by the frequency-tuned saliency detection algorithm can be considered as very satisfactory. The complete samples set consist of 1860 images. We chose two classes of data, high ones containing samples with aesthetics scores above 5.8 and low ones containing samples with aesthetics scores below 4.2. Then the high set contains 434 samples and the low contains 503. As to give equal prior of the two classes, we replicate some of the high set randomly to increase to 503 samples. So the whole training set becomes 1006. Similarly, the amount of all samples goes to 1929.

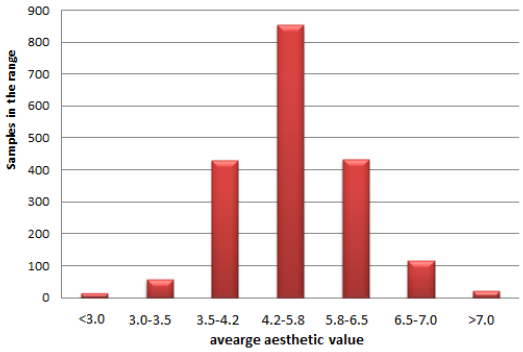

Fig. 3. The distribution of images according to their aesthetic value

4.2 Experimental Results

For the SVM performed on the individual features, the 27 features are obtained by directly extracted from the images. In decreasing prediction accuracy order, the features are { f_7, f_{50}, f_{21}, f_{12}, f_{53}, f_{24}, f_{15}, f_{51}, f_{30}, f_{26}, f_{17}, f_{34}, f_2, f_{49}, f_{22}, f_{13}, f_{52}, f_{46}, f_{18}, f_9, f_6, f_{25}, f_{23}, f_{16}, f_{20}, f_{11}, f_3 }. The correct classification rate achieved by single feature is f_7 with 62.8%. After the run of wrapper based greedy algorithm, features are clearer, they are { f_7, f_{15}, f_6, f_9, f_{11}, f_{20}, f_{12}, f_{18}, f_{46}, f_3, f_{52}, f_{17}, f_2, f_{30}, f_{53} }. The accuracy achieved with the 15 features is 72.0%, with the precision of detecting high class being 85.9% and low class 63.8%. Then we use the same features set and run SVM linear regression prediction onto the training set to have a moderate result of variance of 0.43 with all the samples.

Figure 4 shows part of the CART decision tree with all 53 features. In this figure, the decision nodes are denoted by squares while leaf nodes are denoted by circles. The number of the observations in each node and the splits is shown in the figure. Low nodes with low classification are dark and high nodes with high classification are light. We use 5-fold cross validation to increase the accuracy. The complexity parameter is set to 0.0072 yielding 39 splits. We achieve 73% model accuracy and 60% 5-CV accuracy.

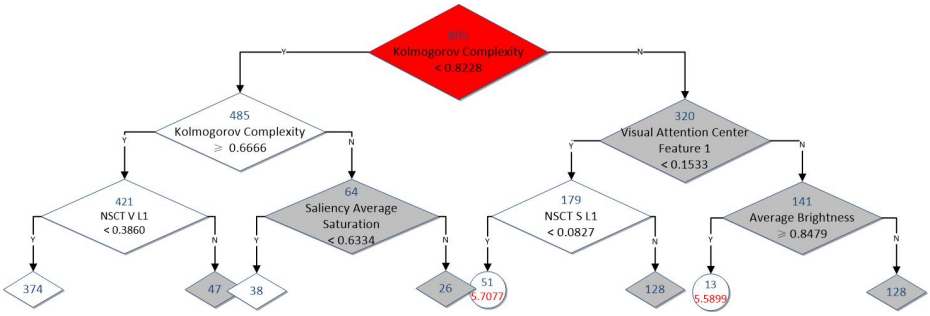

Fig. 4. The CART decision tree obtained

For linear regression, the variance σ^2 of the aesthetics score over 1929 samples is 0.96. With 5 polynomial terms for each of the 53 features, we achieve a residual sum-of-squares R^2_{res} =0.63, which is a 34% reduction of the variance σ^2.

4.3 Discussion

We guess the classification model in the SVM and CART can explain some hidden phenomenon behind the aesthetic evaluation of the viewers. First of all, human eyes and comprehensive system are very vulnerable to complex visual elements. This perhaps because one cannot understand too complex information during a short time or repulsive to incompressible random information as noise. So Kolmogorov complexity largely dominates the classification process from both SVM features filtering process and CART. Next we can see the saliency region of an image plays an important role in determining the image's aesthetic value because half of the features in the SVM selected ones are saliency related. Another essence, we can get from the classification process is that the basic features as brightness, saturation of the images are still prior to human evaluation.

5 Conclusion

We present an automatic approach for machine determining aesthetic value of photographs. We use SVM, CART and linear regression to construct the model. The SVM features filtering is a tedious work, which exploited both filter based and wrapper based method to distill final 15 features for SVM training. The CART is a fast and convenient method (certainly all short enough to within tens of minutes), but it is not so accurate as the former. We also find that the complexity of an image to a great extent dominates the classification process. We also can conclude that human eyes and comprehensive system tend to accept simple and ordered image information. For images with an obvious salient region, viewers are inclined to focus on this region and its features in evaluating the aesthetic value of an image. Furthermore, brightness, saturation etc. are fundamental, but still important factors in aesthetic evaluation.

Acknowledgements. This work is supported by Key Laboratory of Advanced Engineering Surveying of National Administration of Surveying, Mapping and Geo-information (No.TJES1205). We are grateful to the anonymous referees for useful comments and suggestions.

References

1. Flickr, `http://www.flickr.com`
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)

3. Birkhoff, D.G.: Aesthetic measure, Cambridge, Mass. (1933)
4. Machado, P., Cardoso, A.: Computing aesthetics. In: de Oliveira, F.M. (ed.) SBIA 1998. LNCS (LNAI), vol. 1515, pp. 219–228. Springer, Heidelberg (1998)
5. Rigau, J., Feixas, M., Sbert, M.: Conceptualizing Birkhoff's aesthetic measure using Shannon entropy and Kolmogorov complexity. In: Proceedings of the Third Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 105–112. Eurographics Association (2007)
6. Rigau, J., Feixas, M., Sbert, M.: Informational aesthetics measures. IEEE Computer Graphics and Applications 28(2), 24–34 (2008)
7. Romero, J., Machado, P., Carballal, A., Osorio, O.: Aesthetic classification and sorting based on image compression. In: Di Chio, C., et al. (eds.) EvoApplications 2011, Part II. LNCS, vol. 6625, pp. 394–403. Springer, Heidelberg (2011)
8. Datta, R., Wang, J.Z.: ACQUINE: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. In: Proceedings of the International Conference on Multimedia Information Retrieval, pp. 421–424. ACM (2010)
9. Wu, Y., Bauckhage, C., Thurau, C.: The good, the bad, and the ugly: Predicting aesthetic image labels. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 1586–1589. IEEE (2010)
10. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 419–426. IEEE (2006)
11. Wong, L.K., Low, K.L.: Saliency-enhanced image aesthetics class prediction. In: 2009 15th IEEE International Conference on Image Processing (ICIP), pp. 997–1000. IEEE (2009)
12. Luo, Y., Tang, X.: Photo and video quality evaluation: Focusing on the subject. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 386–399. Springer, Heidelberg (2008)
13. Bhattacharya, S., Sukthankar, R., Shah, M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the International Conference on Multimedia, pp. 271–280. ACM (2010)
14. Achanta, R., Hemami, S., Estrada, F.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1597–1604. IEEE (2009)
15. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(12), 2341–2353 (2011)
16. Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2206–2213. IEEE (2011)
17. Cover, T.M., Thomas, J.A.: Elements of information theory. John Wiley & Sons (2012)
18. Da Cunha, A.L., Zhou, J., Do, M.N.: The nonsubsampled contourlet transform: theory, design, and applications. IEEE Transactions on Image Processing 15(10), 3089–3101 (2006)