Detecting Rule of Simplicity from Photos

Long Mai, Hoang Le, Yuzhen Niu
Portland State University
Portland, OR 97207-0751
{mtlong, hoanl, yuzhen}@cs.pdx.edu

Yu-Chi Lai
National Taiwan University of Science and Technology
#43, Sec. 4, Keelung Rd., Taipei 106, Taiwan
yu-chi@mail.ntust.edu.tw

Feng Liu
Portland State University
Portland, OR 97207-0751
fliu@cs.pdx.edu

ABSTRACT

Simplicity refers to one of the most important photography composition rules. Simplicity states that simplifying the image background can draw viewers' attention to the subject of interest in a photograph and help them better comprehend and appreciate it. Understanding whether a photo respects photography rules or not facilitates photo quality assessment. In this paper, we present a method to automatically detect whether a photo is composed according to the rule of simplicity. We design features according to the definition, implementation and effect of the rule. First, we make use of saliency analysis to infer the subject of interest in a photo and measure its compactness. Second, we segment an image into background and foreground and measure the homogeneity within the background as another feature. Third, when looking at an image created with the rule of simplicity, different viewers tend to agree on what the subject of interest is in this photo. We accordingly measure the consistency among various saliency detection results as a feature. We experiment with these features in a range of machine learning methods. Our experiments show that our methods, together with these features, provide an encouraging result in detecting the rule of simplicity in a photo.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications

Keywords

Photo Quality Assessment, Photography Rules Detection

1. INTRODUCTION

Photo composition is an important technique used by photographers to create a high-quality photo. Composition refers to the placement of visual elements in a photo. Among a variety of composition rules, *simplicity* is one of the most important ones. Simplicity is used to make the subject of interest in a photo standing out from its surroundings [10]. Simplicity can be achieved by simplifying the image background, such as making it homogeneous, empty, or completely out of focus, as shown in Figure 1 (a) and (b). A photo that respects the rule of simplicity can easily draw viewers' attention to the subject of interest and help them better comprehend and appreciate it. In contrast, an image with cluttered background is difficult for viewers to focus on the main subject, as shown in Figure 1 (c).

(a) Homogenous or empty background

(b) Out-of-focus background (c) Cluttered background

Figure 1: Rule of simplicity. Simplifying the background in an image by making it homogenous, empty, or out of focus, can emphasize the subject of interest and help viewers to better comprehend and appreciate it (a) and (b). An image with cluttered background is difficult for viewers to focus on the main subject (c).

Computational understanding of how a photo is created is important for applications like photo quality assessment and photo authoring. For example, while the rule of simplicity has been widely practised by professional photographers, amateur users often tend to ignore it or cannot effectively use it due to the lack of expertise or equipment. The existence of the rule of simplicity in a photo is a good evidence of high quality.

This paper focuses on the rule of simplicity and presents a method to detect it in a photo. Detecting the rule of simplicity requires identifying the subject of interest in a photo. While image content understanding has been progressing these years, generic content understanding is still an ongoing research problem. We use saliency analysis as

MM 12, October 29–November 2, 2012, Nara, Japan.

an alternative. Saliency measures low-level stimulus to the human vision system [8] and has been used to infer important content in an image in multimedia applications such as multimedia retargeting [14] and video summarization [13]. We consider that an image with a compact saliency distribution is likely to respect the rule of simplicity. Second, when looking at a photo created with the rule of simplicity, the subject of interest captures a viewer's eye immediately. Therefore, different viewers tend to agree on what the subject of interest is in the photo. We then measure the consistency among various saliency detection results as a feature. Third, we segment an image into background and foreground and measure the homogeneity within the background as another feature. These features are used in a range of machine learning algorithms to detect the rule of simplicity.

Our work is relevant to the recent research on photo quality assessment (e.g. [4, 9, 11, 12]). These methods measure how an image respects the rule of simplicity as a feature and use it together with others for photo quality assessment. This paper aims to explicitly determine whether the rule of simplicity is applied to an image by designing novel features and using them in machine learning algorithms for detection. To the best of our knowledge, our method is the first to detect the simplicity photo composition rule in a photo.

2. FEATURE DESIGN

We design three types of features according to the spatial distribution of visual elements, the background simplicity, and the consistency among the saliency detection results from different methods. We below describe how these features are designed to detect the rule of simplicity in a photo.

2.1 Subject of Interest Compactness

The rule of simplicity recommends that the subject of interest should be surrounded by a simplified background. This suggests that the image region with the subject of interest tends to be compact. We accordingly measure the compactness of the subject of interest in an image as a feature to detect the rule of simplicity.

As inspired by the success of image saliency in multimedia applications, we use saliency analysis to infer the subject of interest. Specifically, saliency analysis produces a saliency map for an input image. Since saliency analysis is often noisy, we divide each saliency map into $n \times n$ blocks with $n = 20$ and compute the average pixel saliency value in a block as the saliency value for the block. We then identify the subject of interest as a minimal number of blocks that altogether contain at least $\alpha\%$ of the total saliency in the image. We finally compute the ratio between the number of blocks in the subject of interest and the total number of blocks in an image to measure the compactness of the subject of interest.

$$f_{cpt} = \frac{N_\alpha}{N} \tag{1}$$

where N_α is the number of blocks in the subject of interest and $N = n^2$ is the total number of blocks in the image.

The performance of this feature relies heavily on saliency analysis. We therefore use the three saliency analysis algorithms that have the highest performance as reported in [2], namely GBVS [6], FT [1], and GC [2]. In addition, the value of α is critical for our method since it affects how well the subject of interest is detected. A big α value can include the whole subject of interest; however, it will include some of the background area. On the other hand, a small α value can make sure that only the image region belonging to the subject of interest is included; however, the selected region may miss some part of the subject of interest. We set the α value experimentally by cross validation on a training data set. The performance of this feature in detecting the rule of simplicity with respect to the α value is illustrated in Figure 2. This figure shows that the optimal α value for the GBVS saliency map, FT saliency map, and GC saliency map are 85%, 70%, and 55% respectively.

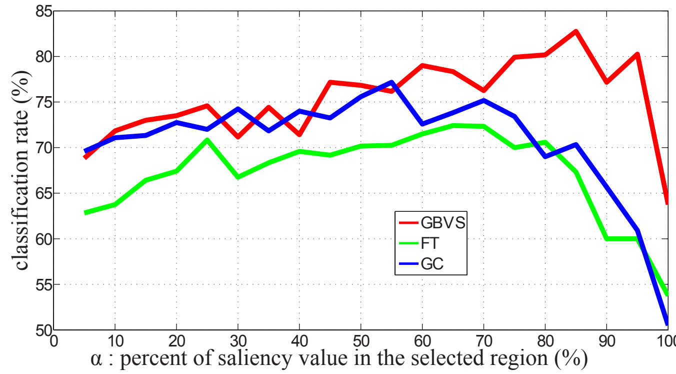

Figure 2: classification accuracy vs. α value

2.2 Background Simplicity

The rule of simplicity states that the background should be simplified. We accordingly measure the simplicity in the background as a feature. We segment an image into two regions using the normalized cut image segmentation method [15] and select the big region as the background.

We measure the simplicity value in the background as the average distance between two image blocks.

$$f_{sim} = \frac{1}{|\mathcal{B}|} \sum_{b_i, b_j \in \mathcal{B}} d(b_i, b_j) \tag{2}$$

where b_i and b_j are two blocks in the background $\mathcal{B}$, and $d(b_i, b_j)$ is the distance between these two blocks. We use the distance function from the segmentation method [15].

$$d(b_i, b_j) = e^{\frac{-||X(b_i)-X(b_j)||^2}{\sigma_s}} e^{\frac{-||F(b_i)-F(b_j)||^2}{\sigma_c}} \quad \text{where } F(b) = [v, v.s.\sin(h), v.s.\cos(h)] \tag{3}$$

where $X(b)$ denotes the position of the block center of b. $F(b)$ is a color descriptor vector, where h, s, and v are the hue, saturation, and value for the average color of b in the HSV color space. σ_s and σ_c are two parameters with value 1.5 and 0.01 respectively.

2.3 Saliency Detection Consistency

The first two features are constructed according to how the rule of simplicity is defined and implemented. We now describe how we define a feature based on the goal and effect of applying the rule. A photo created by following the rule of simplicity has its subject of interest singling out from the surroundings and therefore easily grabs a viewer's attention. Different viewers tend to agree on what the subject of interest is in a photo. As saliency analysis measures how an image element grabs the attention of the human vision system at the early stage of the visual process, we consider

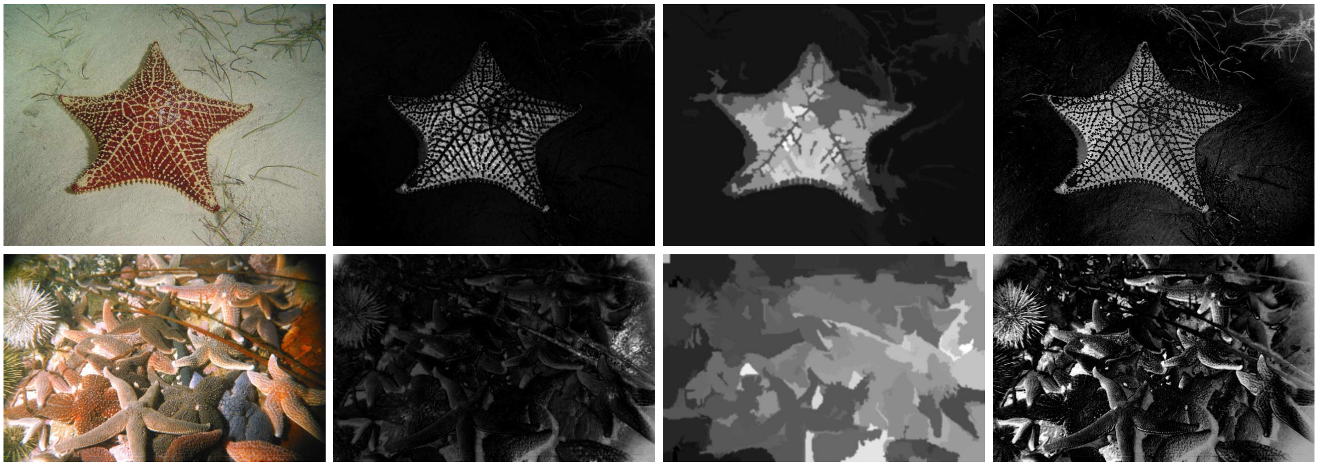

Figure 3: Saliency analysis consistency. The top row shows the saliency analysis results from different methods ([1][2][16]) for a photo that follows the rule of simplicity and the bottom row shows the results for a photo that does not.

that different saliency analysis methods should largely agree on the most salient content in an image if it follows the rule of simplicity. Figure 3 shows the saliency maps for an image that respects the rule of simplicity and an image that does not. We accordingly measure the consistency among saliency maps from different methods as the third feature.

Our method uses seven different saliency analysis methods to estimate seven saliency maps for each image [2, 6, 1, 7, 8, 16][1]. The consistency between every two saliency maps is then computed and concatenated into a feature vector as follows,

$$f_{cnst} = [C(m_i, m_j)]_{i,j \in \{1..7\}, i>j} \tag{4}$$

where $C(m_i, m_j)$ is the consistency value between two saliency maps m_i and m_j.

Computing the consistency value $C(m_i, m_j)$ using all the elements in two saliency maps m_i and m_j is often problematic for two reasons. First, different methods assign saliency values in different scales and simply normalizing a map into the same range is often insufficient. Second, while different saliency methods tend to agree on what the subject of interest is in a photo, they often lead to different saliency analysis results in the background. We solve these two problems by labeling the top 60% image blocks with highest saliency values as salient blocks and then computing the salient region overlap between the two maps.

$$C(m_i, m_j) = \frac{\sum_b \mathbf{1}(b)}{|m_i|} \tag{5}$$

where b is an image block and $|m_i|$ is the number of blocks in m_i. $\mathbf{1}(b)$ is an indicator function that takes value 1 if b is selected as salient in both $m_i(b)$ and $m_j(b)$ and 0 otherwise.

Figure 4 shows the histograms for saliency consistency measurement between saliency maps from the FT [1] and SR [7] saliency detection methods over a set of 400 images that respect the rule of simplicity rule (positive set) and another set of 400 images that do not respect the rule (negative set). The histograms show that the distributions over the positive and negative set for the consistency measurement are clearly different and the consistency values from the positive set are higher than the negative counterpart. This indicates that the feature based on saliency analysis consistency can be used to identify whether an image respects the rule of simplicity or not statistically.

[1]There are two methods in [2].

Figure 4: Consistencies between FT and SR saliency maps over a positive and negative set, respectively.

3. RULE OF SIMPLICITY DETECTION

We use the presented features in a range of classic machine learning algorithms to the rule of simplicity detection, including Logistic Regression, Support Vector Machine (SVM) [3], AdaBoost [5], and K-Nearest-Neighbor method (kNN).

We build a benchmark to evaluate our methods. This benchmark was derived from the photo aesthetics benchmark published by Datta *et al.* [4]. We manually label a set of 1964 photos that respect the rule of simplicity as a positive set, and a set of 2082 photos that do not respect the rule as a negative set. We randomly allocate 70% of the dataset into the training set and the rest 30% into the testing set. For each experiment, we repeat the random partition process 50 times and report the average result.

We first test the effectiveness of each type of feature and their combination for detecting the rule of simplicity using the Logistic Regression classifier. Figure 5 shows the precision-recall curve for each type of feature. All the three features can help achieve a good detection result. The background simplicity alone performs a little worse than the

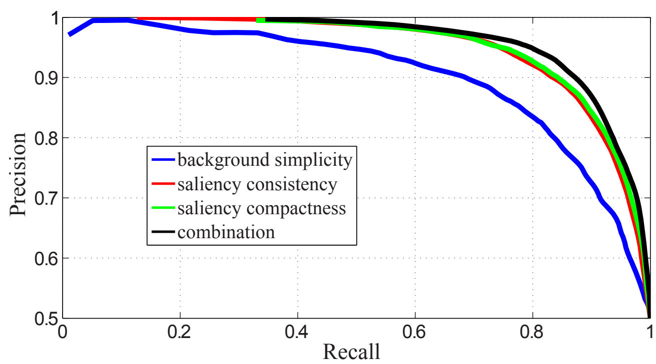

Figure 5: Performance of different features.

other two for mainly two reasons. First, it depends on automatic image segmentation, which is sometimes unreliable. Second, our method currently only uses color to measure the similarity between image blocks in the background. We find that color alone is insufficient as it cannot account for texturness and other type of visual features. Nevertheless, this feature still contributes to the final result when we combine all the three features together. We also find that combining all the three features outperforms each feature individually.

We further report the classification accuracy rates with each individual feature and their combination using these machine learning algorithms in Table 1. Overall, our method achieves around 89% of accuracy in detecting the rule of simplicity in photographs.

Limitations. Our method currently relies on saliency analysis for identifying the subject of interest. Our experiments show that saliency analysis alone is sometimes insufficient and misleads the detection, which contributes to the majority of the failing cases. A better saliency analysis method can improve our method. In the future, we also plan to incorporate object detection in our method. For the background simplicity feature, our method relies on image segmentation to automatically extract the background. Automatic foreground and background segmentation is difficult, which also contributes some of the failure cases.

4. CONCLUSION

In this paper, we presented a method for detecting the rule of simplicity from a photo. We designed features according to the definition, implementation, and effect of the rule, including the saliency compactness, background simplicity, and saliency consistency. We tested these features within a range of classic machine learning algorithms. Our experiments show that our method, together with these features, achieve an encouraging result in detecting the rule of simplicity in photographs. Our research contributes to the computational understanding of photography, which can be used in photo quality assessment and photo composition.

Acknowledgments

The images shown in Figure 1 (a) and (b) are used under a Creative Commons license from Flickr users, including autan, jcoelho, and bonguri. The images shown in Figure 1 (c) and Figure 3 are from Wikimedia Commons, a freely licensed media file repository. Yu-Chi Lai is supported by NSC-100-2221-E-011-151.

	Regression	kNN	SVM	AdaBoost
saliency compactness	86.7%	86.7%	87.6%	87.4%
background simplicity	82.2%	82.1%	82.8%	82.1%
saliency consistency	86.0%	85.7%	87.8%	86.4%
all	87.7%	86.8%	89.2%	88.5%

Table 1: Simplicity rule detection accuracy with various features and machine learning algorithms

5. REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1597 –1604, 2009.

[2] M.-M. Cheng, G.-X. Zhang, N. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 409–416, 2011.

[3] C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, pages 288–301, 2006.

[5] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.

[6] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in Neural Information Processing Systems*, pages 545–552, 2007.

[7] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[8] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[9] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 419–426, 2006.

[10] B. P. Krages. *The Art of Composition*. Allworth Communications, Inc., 2005.

[11] W. Luo, X. Wang, and X. Tang. Content-based photo quality assessment. In *IEEE International Conference on Computer Vision*, pages 2206 –2213, 2011.

[12] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*, pages 386–399, 2008.

[13] Y. Ma, L. Lu, H. Zhang, and M. Li. A user attention model for video summarization. In *ACM Multimedia 2002*, pages 533–542, 2002.

[14] Y. Niu, F. Liu, X. Li, and M. Gleicher. Warp propagation for video resizing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 537–544, 2010.

[15] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[16] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006.