Scoring Photographic Rule of Thirds in a Large MIRFLICKR Dataset: A Showdown Between Machine Perception and Human Perception of Image Aesthetics

Adnan Firoze(✉), Tousif Osman, Shahreen Shahjahan Psyche, and Rashedur M. Rahman

Department of Electrical and Computer Engineering, North South University, Plot-15, Block-B, Bashundhara Residential Area, Dhaka, Bangladesh
{adnan.firoze, tousif.osman, shahreen.psyche, rashedur.rahman}@northsouth.edu

Abstract. In this research we have developed and evaluated a system that uses the image compositional metric called 'Rule of Thirds' used by photographers to grade visual aesthetics of an image. The novel aspect of the work is that it combines quantitative and qualitative aspects of research by taking human psychology into account. The core idea is to identify how similar the perception of a 'good image' and 'bad image' is by machines versus humans (through a user study based on 255 participants on 5000 images from the standard MIR-FLICKR database [9]). We have considered the compositional norm, namely 'rule of thirds' used by photographers and inspired by the golden ratio that states that - if an image is segmented on a 3×3 grid, then it is appealing to the eye when the most salient object(s) or 'subject(s)' of the image is located precisely on or aligned on the middle grid lines [11]. First, we preprocess the input image by labeling the regions of attraction for human eye using two saliency algorithms namely Graph-Based Visual Saliency (GBVS) [3] and Itti-Koch [4]. Next, we quantify the rule of thirds property in images by mathematically considering the location of salient region(s) adhering to rule of thirds. This is then used to rank or score an input image. To validate, we conducted a user study where 255 human subjects ranked the images and compared our algorithmic results, making it a both a quantitative and qualitative research. We have also analyzed and presented the performance differences between two saliency algorithms and presented ROC plots along with similarity quantification between algorithms and human subjects. Our massive user study and experimental results provides the evidence of modern machine's ability to mimic human-like behavior. Along with it, results computationally prove significance of rule of thirds.

Keywords: Image processing · Visual perception · Image saliency Rule of thirds · Photography · Golden ratio · Computer vision Image score · Image composition · Image · Flickr

N. T. Nguyen et al. (Eds.): ACIIDS 2018, LNAI 10751, pp. 466–475, 2018.
https://doi.org/10.1007/978-3-319-75417-8_44

1 Introduction

One of the most abstract instance of comprehending visual data is 'finding beauty' - only possible by human psychology. In this research, we have devised a synthetic system which can understand visual appeal in a framed image. We have used the basic photographic rules as the foundation and the rules of computer vision and machine learning as the pillars of our system. In this research, first we have identified the salient region of an image using methods derived from research of Itti, Koch, and Harel [3, 4]. Next, we applied photographers' rule of thumb to measure the aesthetic appeal of the image [11]. The metric we have narrowed down to is 'Rule of Thirds' (ROT). Finally, we conducted a survey to evaluate our system's aesthetic measurement of an image with respect to visual aesthetics perceived by human psychology.

2 Related Works

Few have traveled before us in finding automated means to extract compositional properties in images. Here, the novel idea is the human user-study that we conducted. We have incorporated ideas, findings, and bits and pieces of several research in our system to conduct our research. One of these research aims to detect ROT compositions in images [1]. The ROT states that placing important or salient objects along the images' thirds lines (refer to Fig. 1) or around their intersections often produces highly aesthetic photos [2]. In this research, researchers have utilized multiple image saliency algorithms, namely, Fourier Transform (FT) Map, Graph Based Visual Saliency (GBVS) map [3], Global Contrast (GC) map and Objectness (OBJ) map to extract the salient region of an image. Afterwards, several machine learning techniques including the Naïve Bayesian Classifier, Support Vector Machine (SVM) etc. have been used for the rule of thirds detection. This research also shows that GBVS performs the best exhibiting 75% accuracy over a dataset of 2089 images collected from Flickr and Photo.net. We have incorporated the ideas of finding salient regions and ROT compositions of this research into our system.

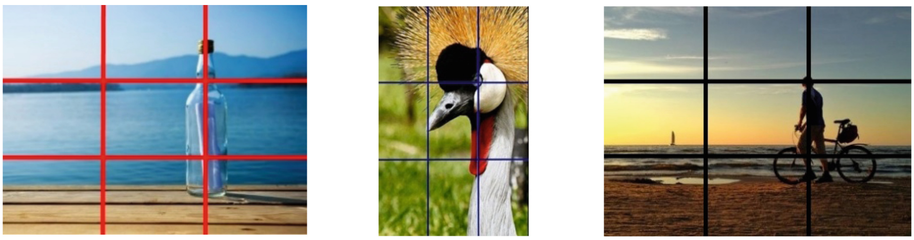


Fig. 1. Some sample images that follows rule of thirds in photography to enhance appeal or aesthetics

In research [3], authors have proposed a bottom-up saliency model named “Graph Based Visual Saliency” (GBVS). This model mainly works in two steps. While the classical algorithms of Itti and Koch achieved only 84% of the ROC area of human based control, GBVS has achieved 98%. In a prior research [4] by Itti and Koch, authors aimed to solve this same problem of finding salient regions in digital images.

In [5] Mai et al. has used Rule of Thirds composition method to measure the aesthetics of an image. To do that, first they have used a variety of saliency and generic objectness methods to detect the main object of an image and after that they have used the concept of Rule of Thirds to measure the aesthetics of that particular image. For detecting a salient region of an image, in this paper, they have mainly used three algorithms and they are GBVS (Graph-based Visual Saliency), FT (Frequency-tuned salient region) and GC (Global contrast based salient region detection). Furthermore, they have used generic objectness analysis as a complement of saliency analysis. For the detection of ROT they have used various Machine Learning methods, for example, Naïve Bayesian Classifier, Adaboost etc.

Amirshahi et al. [6] have contrasted aesthetically pleasantness of photographs and paintings between computer based scoring and behavioral scoring (scores that are obtained from 30 participants) on the basis of Rule of Thirds compositional method.

Maleš et al. in their paper [7], they have presented a saliency based method to detect the compositional rule – ROT. To detect a salient region they have used two algorithms – Context Aware (CA) salient region detector and Global Contrast (GC) based salient region detector. After that, they have created a training set on which they have applied Principal Component Analysis (PCA). They have used Linear Discriminant Analysis, Mahalanobis Linear Discriminant Analysis, Quadratic Discriminant Analysis and Support Vector Machines to train the classifiers.

In [8], the authors have presented a collection for MIR community which comprises a subset of 25,000 images from the Flickr website under creative commons license that had been a standard dataset for most research in visual data retrieval. These images are also redistributable for research purposes and, also represent a real user community.

3 System Design

Our system takes images as input and produces a score as aesthetic appeal. As mentioned in the abstract and introduction, we have used metric - ROT in our system to produce the beauty measurement. We have accomplished the task of scoring beauty in several stages starting from an input image to final beauty measure. Therefore, we have subdivided the processing of the system into stages presented in Fig. 2. Figure 2 represents simple processing flow of our system. In the initial stage, the system receives image as input from the users via the user interface. Next, that image is fed into the preprocessor which marks the beginning of the second stage of our system. To apply the compositional metric, precondition is to compute the salient regions of the image. Hence the second stage has been developed using GBVS and Itti-Koch saliency algorithms to calculate salient regions.

Based on user selected algorithm out of two (Itti-Koch and GBVS), salient region(s) of the input image are calculated and are passed to the third stage. This stage is also the

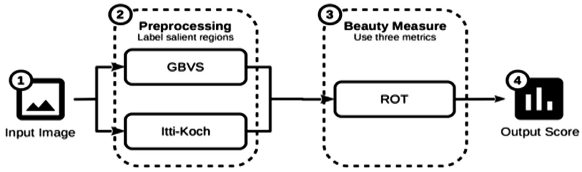


Fig. 2. Simple block diagram of processing stages.

centerpiece of our system. This stage of processing has an engine running as a service in the system, which receives the input image and salient image map from previous stage. Based on the ROT evaluator engine, the system processes it to produce ROT measure. This processing stage is the heart of this research.

Inner workings of this engine will be explained in the following subsections. Finally, in the fourth and final stage the beauty score is transferred to a user interface from the service for representing the output to the users of the system. We have developed a system consisting a web interface powered by Node.JS and an image processing service for ROT metric developed using MATLAB. As a result, users can access this tool just using a web browser without needing access to MATLAB or any other additional software. The web server and the MATLAB engine communicate via JSON. Figure 3 represents a simple overview of our system design.

Fig. 3. Simple overview of system design

For validation, we have conducted a user study to evaluate our system's aesthetic pleasantness measurements against visual aesthetics perceived by humans. We have executed our system on 5000 images from MIR Flickr dataset [11] and analyzed the execution and performance differences between GBVS and Itti-Koch algorithm.

3.1 Workflow of the ROT Engine

Definition of Rule of Thirds or ROT states that in a rectangular frame, human eye tends to perceive objects more appealing - the more they are close to the gridlines having 2/3 area on the greater side and 1/3 area on the shorter side [2]. This visual compositional rule has been established based on the golden ratio [9] proportion guideline by the ancient Greeks [2]. The ROT was first documented and written down in 1797, in the book – 'Remarks on Rural Scenery' by J.T. Smith [10]. In simple terms, if we draw two evenly spaced vertical and two evenly spaced horizontal imaginary gridlines in a rectangle image then each of these lines divides the image according to golden ration. Figure 4 simplifies the understanding of the subdivision.

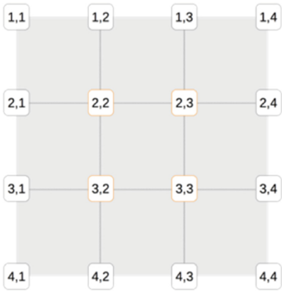


Fig. 4. Corner points of gridlines.

An image having its salient regions closer to any of these lines will have better aesthetic appeal. Furthermore, salient objects closer to the intersecting points of the gridlines will be more appealing to human eye as it means the object is maintaining golden ratio vertically and horizontally both at the same time. To measure distance, we have calculated the center point or centroid ***C (x, y)*** of salient area in the image. Then the distance is calculated with respect to the centroid. Centroid is calculated as follows for a given salient blob:

$$C_x = \frac{X_1 + X_2 + X_3 + \ldots + X_k}{S} \tag{1}$$

$$C_y = \frac{Y_1 + Y_2 + Y_3 + \ldots + Y_k}{S} \tag{2}$$

Where, X_1, X_2, X_3, …. X_k are the x coordinates, Y_1, Y_2, Y_3, … Y_k are the y coordinates, and S is the sum of all pixels in the salient blob. We have constructed a distance function (Eq. 3) in this research which calculates a distance measurement from each cross sections of gridlines and the centroid of the salient region of the image. This measurement is the proxy of measuring beauty and visual aesthetics.

3.2 Distance that Serves as Scoring Function

To calculate distance, we have measured the distance of centroid from 2/3 cross-sections of each gridline. This gives us a tentative distance measurement from individual gridlines. Pair with the shortest measurement is the candidate axis for further calculation as the salient region would be the closest with that line. An image can have n number of salient regions where n is a positive integer. Therefore, the input image can have n number of distance scores for each subject or salient region. Regarding understandability and writing progression we have used ROI (Region of Interest) as synonym of salient regions in the next few sections. In Fig. 4, (1, 1), (1, 2), (1, 3), …, (4, 4) are the corner points of gridlines. If g (2, 2), g (2, 3), g (3, 2), g (3, 3) are the cross-sectional points then the distance is measured as Eq. 3.

$$distance_{ROI}(n) = \min\Big(\sqrt{(G_{22x} - C_x)^2 + (G_{22y} - C_y)^2}, \sqrt{(G_{23x} - C_x)^2 + (G_{23y} - C_y)^2}, \sqrt{(G_{33x} - C_x)^2 + (G_{33y} - C_y)^2}, \sqrt{(G_{32x} - C_x)^2 + (G_{32y} - C_y)^2}\Big) \tag{3}$$

3.3 Scoring Function Normalization

As an input image can have more than one ROI, hence it can have more than one distance score. The distance is a measurement in pixels with respect to individual image (where size or dimensions can vary) which is not an absolute measurement that we can compare with scores of other images. Also, the more the Region of Interest (ROI) is closer to prominent gridline, less the distance measurement but aesthetic score is the compliment. Considering these facts, we have constructed the following equation to normalize distance score and produce a score which we can considered as the aesthetic score of an image.

$$score(img) = \left(\sum_{i=1}^{all\,ROIs=N}\left(\frac{Area\ of\ ROI_i \times 0.05}{(img.length \times \mathrm{img}.width)} \times \left(1 - \frac{distance_{ROI}(i)}{\sqrt{img.height^2 + img.width^2}}\right)\right)\right) \times \frac{1}{N} \tag{4}$$

In Eq. 4, distance has been divided by the diagonal of the image to produce a relative measurement and which is then subtracted form one. This measurement is then multiplied with the relative area of ROI to give better score to ROI having bigger area and negate ROIs with smaller areas. Finally scores of all ROIs has been added and divided by the number of ROIs to compute the final aesthetic score.

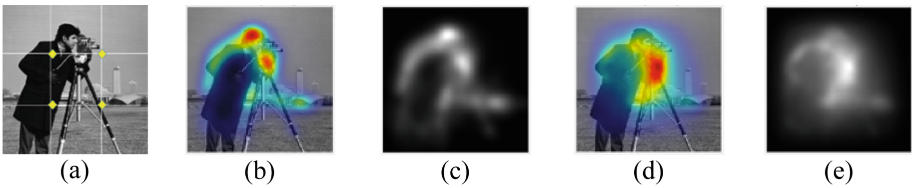
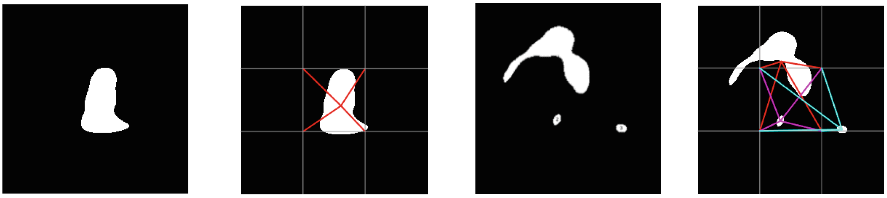


Fig. 5. GBVS vs. Itti-Koch. (a) cameraman.tif image on grid, (b) Itti-Koch Map Overlayed, (c) Itti-Koch Map and (d) GBVS Map Overlayed, and (e) GBVS Map. (Color figure online)

For illustration, we decided to use the standard “cameraman.tif” file (Fig. 5) and produced a running example. Beforehand, if we look closely we can observe this image is not especially good in terms of ROT. In the first step of processing we draw imaginary gridlines over the image. In Fig. 5a, white lines are the gridlines and the yellow dots are the cross-sections of gridlines. Next, we use GBVS (Fig. 5d, e) and Itti-Koch maps (Fig. 5b, c) to produce ROI for the input image maps. Observe that there are one ROI for GBVS and three ROIs for Itti-Koch algorithm. After that, the system uses distance function and next our normalization technique produces final aesthetic score.

From Fig. 6(a) we can see that, we have gotten only one ROI using GBVS algorithm (from ‘cameraman.tif’). So our whole procedure is developed surrounding only this ROI. Firstly, we have measured the centroid of the ROI and the axis of the centroid in pixel is ($C_{x,}$ C_y) = (197px, 154px). After that, using Eq. 3 we have calculated the minimum distance between the gridlines and the centroid which is 53.96px. The area of our ROI is roughly 7146px^2 and the height and width of the image are 311px and 376px accordingly. So, finally using Eq. 4 we got the final score for this particular image that is 0.55 (where the scale is as such that 1 is the best and 0 being the least aesthetic).



(a) (b) (c) (d)

Fig. 6. ROI and distance measurement using GBVS and Itti-Koch. (a) ROI using GBVS, (b) Distance score using GBVS (c) ROI using Itti-Koch, and (d) Distance score using Itti-Koch.

4 User Study

We have conducted a user study to evaluate our system. We used 255 human participants in the study and they were aged between 19 to 29 regardless of gender. They were asked to score randomly presented images. All participants were undergraduate students from the North South university and some of the ware member of photography club. Each participant was presented with 35 images from different category in random order to eliminate bias. Participants were asked the question – ***“Rate the picture you are seeing, where 5 is the best score 1 is the worst score. Also, click on the most salient object or region you think is present in this image using Javascipt.”*** We recorded both ratings and the pixels of the participants perceived ‘subject’ of the image.

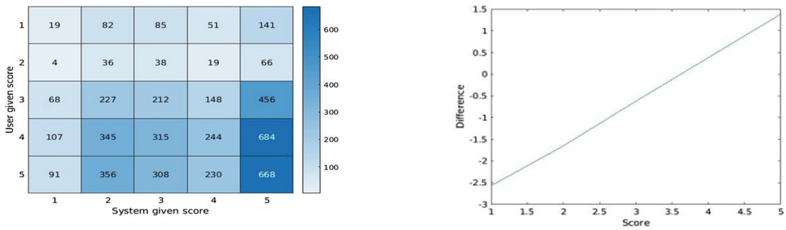
We have used 5000 images from this dataset to evaluate our system.

5 Results Analysis

In this research, we have two results and analyzed their accuracy. Firstly, we produced aesthetic score which we measured against aesthetic score response to human psychology, and then, we compared the execution and performance differences of the two saliency algorithms we have used in this research. Our results and analysis have been described in detail in the following subsections.

5.1 Distances from Baseline of Human Users

We have presented the results in Fig. 7. In Fig. 7a, the horizontal axis represents the system given score and the vertical axis represents the user given score. Both of the scores are given in a scale between 1 to 5, and the values in each box represent the number of responses against each score for both system and human. For instance, in position (5,5) the number is 668 means that 668 images were scored 5 by the both system and user. Figure 7b represents score vs. the score difference between the user user and the system. From this figure we can observe, for lower score i.e. 1, the average difference is higher – 2.5 and for higher score i.e. 5, the average difference is lower – 1 (note: lower the better). This gives us the intuition of the fact that in most cases when humans find some image appealing it maintains minimum level of ROT but the opposite is not true.



(a) (b)

Fig. 7. (a) System generated scores vs. user generated scores and (b) the differences between the score responses.

The average sum of squared difference (SSD) is used to measure distance of score and provide an analytical result as presented in Eq. 5. Here, *i* is image index (1 to 100) and j is category index (1 to 10).

$$SSD_{ij} = \frac{\sqrt{\sum_{i=1}^{for\ all\ in\ j} \left(score_i - score_{human\ user}\right)^2}}{100} \tag{5}$$

5.2 Comparative Analysis of Saliency Algorithms

To do this comparative analysis of saliency algorithms, first we observed that- although the average SSD is low across each category, some results are too far or too close to the baseline. Bounded boxes were drawn around the detected salient regions for the two algorithms- GBVS and Itti-Koch. After that, we went through all of them manually to further validate the system if the saliency algorithms indeed find the salient regions of an image.

5.3 Methodology of the Comparisons

All of the 5000 images were annotated. Parameters that were noted are – salient regions detected by human, GBVS and Itti-Koch Saliency algorithm. We used human perception to detect four scenarios – true positives (properly detected ROI), false positive (any algorithm that detected a salient region which is not ROI), true negatives (non-salient region detected as salient region) and finally false negatives (salient regions missed by the algorithms). If a salient region detected by any of the three algorithms overlapped 50% or more than that of what human perceived as salient region is recognized as true positive. After that, we represented the research findings using ROC plots. They are presented in Fig. 8. The number of regions of interest or ROI's is 2680 which is higher than 5000 (the number of images) is because many single images contained more or less than one salient regions or subjects.

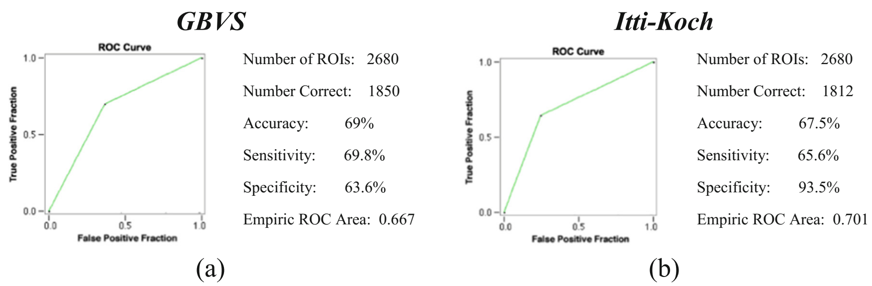


Fig. 8. Comparative ROC plots of between GBVS (a) and Itti-Koch (b) saliency algorithms

6 Discussion and Future Work

In this research, we have successfully devised a system that can perceive beauty with good accuracy. In this regard, we have been successful in most cases. We have analyzed significance of the compositional rule and its combination in different scenarios. We also analyzed differences between GBVS and Itti-Koch saliency algorithm based on real world applications by involving actual human comprehension. Finally, we can conclude that image compositional metrics are a valid and stable mean to understand visual aesthetics and ROT has significance in good magnitude. However, we experimentally validated that GBVS outperforms Itti-Koch to some extent.

Although, we have successfully gone to the closest points with some categories of images, we feel if we use other compositional metrics i.e. Rule of frames, Rule of odds, Rule of space etc. [11] in addition, it would give more insights. A larger user study is coming up as well.

Acknowledgments. This work was funded by North South University's annual research grant for the fiscal year 2017–18. We would like to thank Professor John Kender (http://www.cs.columbia.edu/~jrk/), Professor of Computer Science at Columbia University for his insights.

References

1. Mai, L., Le, H., Niu, Y., Liu, F.: Rule of thirds detection from photograph. In: Proceedings of the IEEE International Symposium on Multimedia (ISM), Dana Point, CA, USA, pp. 91–96 (2011)
2. Peterson, B.: Learning to See Creatively, 1st edn, pp. 92–93. Amphoto Books, New York (2003)
3. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2006)
4. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. **40**(10–12), 1489–1506 (2000)
5. Mai, L., Le, H., Niu, Y., Liu, F.: Rule of thirds detection from photograph. In: IEEE International Symposium on Multimedia, Portland (2011)
6. Amirshahi, S.A., Hayn-Leichsenring, G.U., Denzler, J., Redies, C.: Evaluating the rule of thirds in photographs and paintings. Art Percept. **2**, 163–182 (2014)
7. Maleš, M., Heđi, A., Grgić, M.: Compositional rule of thirds detection. In: IEEE International Symposium on ELMAR 2011, Dana Point (2011)
8. Murillo, A., Košecká, J., Guerrero, J., Sagüés, C.: Visual door detection integrating appearance and shape cues. Rob. Auton. Syst. **56**(6), 512–521 (2008)
9. Huiskes, M.J., Lew, M.S.: The MIR flickr retrieval evaluation. In: ACM International Conference on Multimedia Information Retrieval (MIR 2008), Vancouver, Canada (2008)
10. Weisstein, E.W.: Golden Ratio. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GoldenRatio.html
11. McCurry, S.: 9 Photo Composition Tips (feat. Steve McCurry). Youtube (2015). https://youtu.be/7ZVyNjKSr0M. Accessed 2 March 2016