# Rule of Thirds Detection from Photograph

Long Mai, Hoang Le, Yuzhen Niu, and Feng Liu
*Department of Computer Science*
*Portland State University*
*Portland, OR, USA*
{*mtlong, hoanl, yuzhen, fliu*}*@cs.pdx.edu*

***Abstract*—The *rule of thirds* is one of the most important composition rules used by photographers to create high-quality photos. The rule of thirds states that placing important objects along the imagery thirds lines or around their intersections often produces highly aesthetic photos. In this paper, we present a method to automatically determine whether a photo respects the rule of thirds. Detecting the rule of thirds from a photo requires semantic content understanding to locate important objects, which is beyond the state of the art. This paper makes use of the recent saliency and generic objectness analysis as an alternative and accordingly designs a range of features. Our experiment with a variety of saliency and generic objectness methods shows that an encouraging performance can be achieved in detecting the rule of thirds from photos.**

***Keywords*-Multimedia Quality, Photo Composition, Rule of Thirds**

## I. Introduction

Composition is an important aspect of photo quality. Photo composition refers to the placement of visual elements. For example, professional photographers often divide an image using the imagery horizontal and vertical thirds lines and place important objects along these lines or their intersections, as shown in Figure 1. This particular visual element placement is called *rule of thirds* [1]. The placement that respects the rule of thirds often leads to more visually appealing photos than simply placing objects in the photo center. Photo editing algorithms [2] and tools, like *Adobe Photoshop CS5*, can help meeting the rule of thirds by cropping or warping a photo.

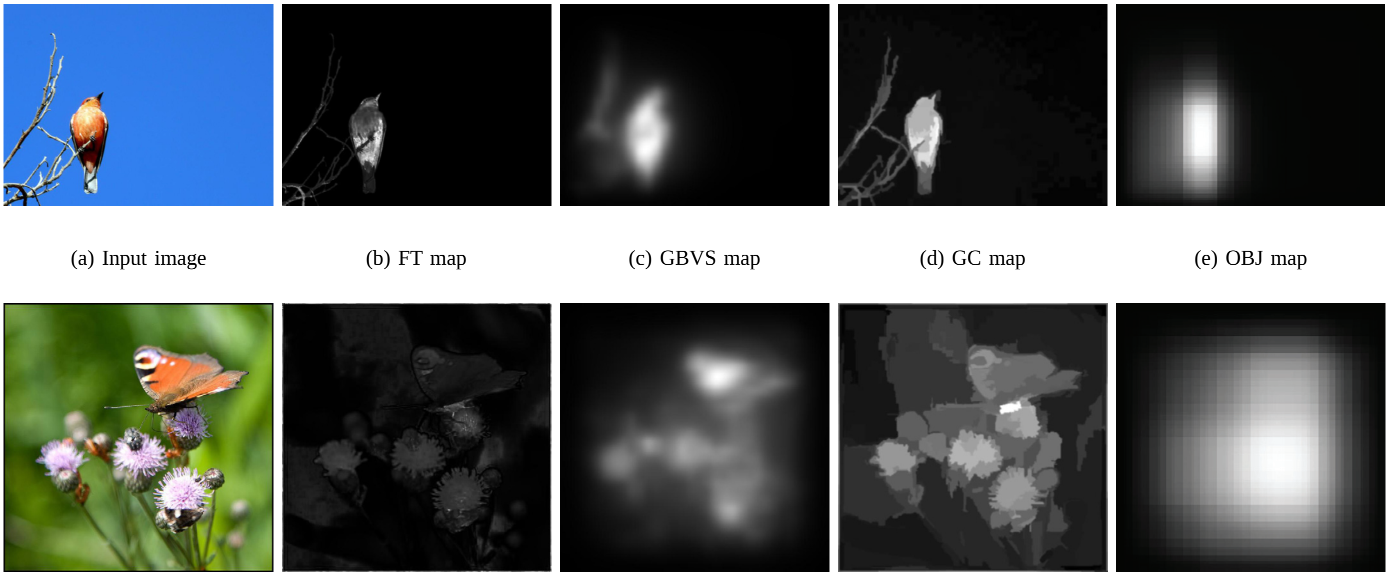
Detecting photo composition is important for measuring photo quality [3], [4]. In this paper, we focus on detecting the rule of thirds from a photo. Ideally, such a detection method needs to know what are important and where they are. This requires the semantic understanding of a photo. While exciting results have been reported on detecting some specific objects, such as face, important object detection, in general, is still a challenging problem. Our idea is to make use of the recent saliency (c.f. [5], [6], [7], [8], [9]) and generic objectness analysis [10] as an alternative to semantic content understanding. While saliency analysis as an alternative to importance detection has been shown successful in some multimedia applications, such as multimedia retargeting (c.f. [11], [12]), it is unclear how they will perform in our problem of the rule of thirds detection. There

Figure 1. Examples of rule of thirds. Professional photographers often place important objects along the thirds lines or their intersections to create visually appealing photos.

is often a gap between the low-level saliency analysis and high-level important content detection.

In this paper, we explore the recent saliency analysis and objectness analysis methods for photography rules of thirds detection. We design a variety of features based on the saliency and objectness map to detect visual elements and infer the spatial relationship among them. We adopt a range of machine learning algorithms for the photography rule of thirds detection using these features. Our experiments with these features show that our method achieves an encouraging detection result and using saliency and generic objectness analysis to detect photo composition rules is promising.

In the rest of the paper, we first describe how we design features based on saliency and objectness analysis in Section II. We then explain how we use these features in a range of machine learning techniques for the rule of thirds



(a) Input image (b) FT map (c) GBVS map (d) GC map (e) OBJ map

(f) Input image (g) FT map (h) GBVS map (i) GC map (j) OBJ map

Figure 2. Image saliency map computed using different methods, including FT [8], GBVS [6], and GC [9], as well as objectness map (OBJ) [10]. These examples show that saliency analysis can often be used to infer the important content location. Sometimes, the saliency analysis results are misleading, such as (g) and (i).

detection and report the results in Section III. We finally conclude this paper in Section IV.

## II. Saliency-based Feature Design

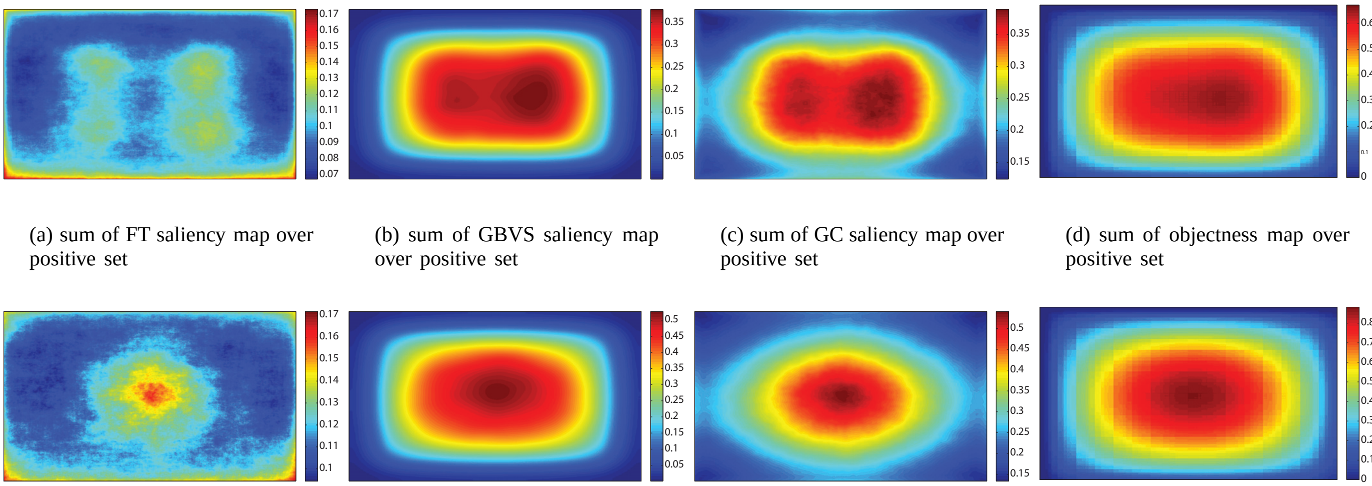
Rule of thirds detection requires the knowledge about important content locations. However, important content detection is beyond the state of the art. Inspired by the success of using low-level saliency as alternative in other multimedia applications, such as multimedia retargeting [12], our method explores the saliency analysis to detect important content. Visual saliency measures the low-level stimuli to the human visual system. A variety of methods have been recently developed to estimate the visual saliency from images. Our method selects three recent algorithms for saliency estimation as successful salient object detection results are reported using these methods. These three algorithms are GBVS [6], FT [8], and GC [9]. Each of these methods takes an image as input and outputs a map which indicates the saliency value at each pixel/block. Figure 2 shows several examples. We omit the description of these methods. Please refer to their original papers for the detail.

To quickly examine whether saliency analysis is useful for our task, we sum up the saliency map for a positive photo collection with 1000 images that respect the rule of thirds and a negative photo collection with 1000 images that do not respect the rule of thirds. We show the summed saliency map for each method in Figure 3. This figure shows that for the summed saliency map for the positive collection, the most salient regions are around the intersections of the thirds lines, no matter which saliency method we use. For the summed saliency map for the negative collection, the most salient regions are around the image center. This shows that saliency map could be potentially useful for the task of detecting the rule of thirds.

As also revealed by the multimedia retargeting research, saliency alone sometimes leads to undesirable results due to the gap between the low-level stimuli analysis and the high-level semantic understanding. This suggests that there may be necessarily a limit on the performance of using saliency for the rule of thirds detection. We design a variety of features based on saliency analysis to examine this limit and best use them for the rule of thirds detection. Further more, our method uses generic objectness analysis as a complement to saliency analysis [10]. This objectness method returns a large number of windows that likely contains an object. Each window comes with a confidence that it contains an object. Our method sums these windows weighted by their confidence values into an objectness map. Each element of this objectness map also indicates the likelihood that it belongs to an object, which is very similar to how our method uses the saliency map to infer the object location. So our method uses the objectness map in the same way as the saliency maps. For simplicity, we refer to this objectness map as a special type of saliency map below.

### A. Saliency Map Centroid

A photo that respects the rule of thirds places important visual elements around the thirds lines or their intersections.



(a) sum of FT saliency map over positive set
(b) sum of GBVS saliency map over positive set
(c) sum of GC saliency map over positive set
(d) sum of objectness map over positive set
(e) sum of FT saliency map over negative set
(f) sum of GBVS saliency map over negative set
(g) sum of GC saliency map over negative set
(h) sum of objectness map over negative set

Figure 3. Sum of saliency map over positive and negative image set, respectively.

Since our method uses saliency analysis to infer important content, we compute the centroid of the saliency map to approximate the important object location. However, the centroid of the whole saliency map or a very large window of the saliency map is often off the important salient region of the image, as shown in Figure 4. Instead of using the centroid of the whole saliency map, our method finds a minimal rectangle that contains at least $\lambda$ of the total saliency value as follows.

$$\mathbf{p}_c = \frac{\sum_{\mathbf{p}\in\mathcal{W}} w_p * \mathbf{p}}{\sum_{\mathbf{p}\in\mathcal{W}} w_p} \tag{1}$$

where $\mathbf{p}_c$ is a two-element vector denoting the centroid location and $\mathbf{p}$ is a two-element vector denoting a point in the minimal rectangle $\mathcal{W}$ that contains $\lambda$ of the total saliency values in the image. $w_p$ is the saliency value at $\mathbf{p}$. Our method uses the summed area table algorithm to expedite the search for the minimal rectangle $\mathcal{W}$ so that the amount of saliency in each window can be computed in constant time [13].

$\lambda$ is an important parameter. On one hand, when it approaches one, $\mathbf{p}_c$ becomes the centroid of the whole saliency map. As shown early, $\mathbf{p}_c$ is off the important region of the image. On the other hand, when $\lambda$ is very small, $\mathbf{p}_c$ often suffers from noise. We set the $\lambda$ value experimentally by cross validation on a training data set. The performance of this feature with respect to the $\lambda$ value is illustrated in Figure 4 (c). This figure shows that the optimal $\lambda$ values for the GBVS saliency map, FT saliency map, GC saliency map, and objectness map are 60%, 30%, 30%, and 40%, respectively.

### B. Saliency Around Thirds Lines and Their Intersections

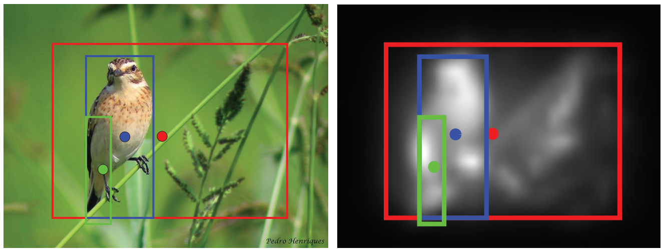
A photo that respects the rule of thirds usually has more saliency around the thirds lines and their intersections. Based on this observation, our method divides an image into a $5\times5$ grid mesh. This grid mesh is built in a way such that it aligns well with the thirds lines. Specifically, we create a strip centered at each thirds line with a width 1/6 of the input image size and split the image into 5 regions in each dimension, as shown in Figure 5. We call this grid mesh a *thirds map*. The average saliency value at each grid cell is computed as a feature.

$$w_i = \frac{\sum_{\mathbf{p}\in\mathcal{W}_i} w_p}{A(\mathcal{W}_i)}, 1 \leq i \leq 5 \tag{2}$$

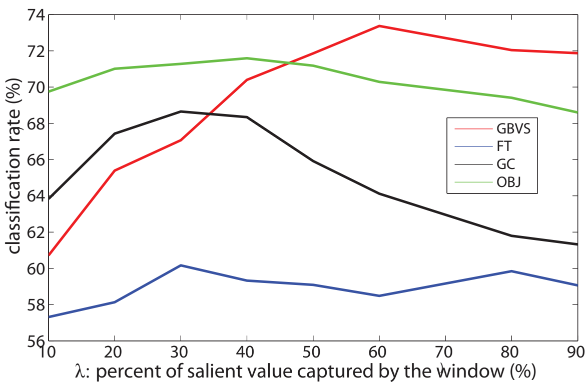
where $A(\mathcal{W}_i)$ is the area of the $i^{th}$ grid cell $\mathcal{W}_i$.

### C. Raw Saliency Map

The saliency map centroid and the third map are designed based on the domain knowledge of the rule of thirds. It is also interesting to examine using the raw saliency map itself for the rule of thirds detection. Specifically, our method resizes each saliency map into a $n \times n$ saliency map and uses this re-sampled saliency map as a feature vector. In our method, we use $n = 20$. Our experiment shows that our method is pretty stable with regard to $n$. Our method further applies the Principal Component Analysis (PCA) to reduce the dimensionality of the feature vector [14]. The number of eigen-vectors used in our method is 15, which is determined experimentally via cross-validation. This eigen analysis of the raw saliency map gives another feature vector that has 15 elements. Figure 6 shows the first 5 eigen-saliency maps in the order of eigen-values. These eigen-vectors show the features of images that respect rule of thirds or not. For



(a) Input image (b) Saliency map




(c) classification accuracy vs. $\lambda$ value

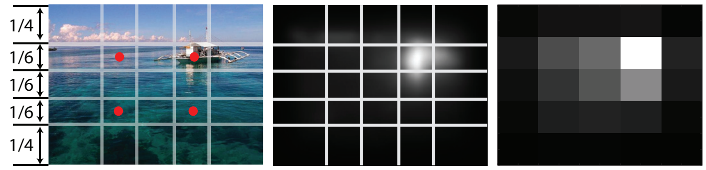
Figure 4. When the centroid is computed from a very big region that contains almost all the saliency values, it is often off the important content location as indicated by the red rectangle and point. On the other hand, when a window that contains a very small amount of saliency is used, its centroid is not an accurate predication of the important content location either, as indicated by the blue rectangle and point. A window that contains an appropriate amount of saliency value is important for inferring the important content location.

example, the first eigen-vector highlights the image center, which intuitively shows a fundamental difference between rule of thirds images and non-rule of thirds images. The other eigen-vectors highlights either the thirds lines or their intersections.

In summary, our method extracts three types of features as described in the above subsections, namely the saliency map centroid, the thirds map, and the eigen coefficients of saliency map. These three type of features are extracted from each of the four saliency maps. So totally our method extracts an overall feature vector with $4 \times 42$ elements for each image.

## III. Rule of Thirds Detection

Our method applies a range of classic machine learning techniques for the rule of thirds detection, including the Naïve Bayesian Classifier, Support Vector Machine (SVM) [15], Adaboost [16], and K-Nearest Neighbor method (kNN). For SVM, our method uses the RBF kernel. We use the LIBSVM [17] implementation. For Adaboost, our method uses the Logist Boost method. We use the OpenCV implementation of Adaboost[1]. For kNN, our method selects k=30 and uses the Euclidean distance as the distance metric.




(a) Input image (b) Saliency map (c) Third map

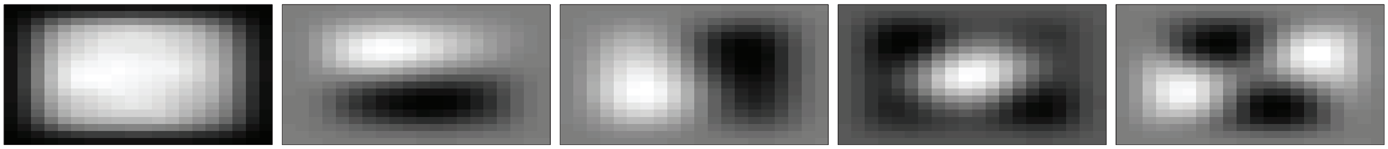
Figure 5. Third map. We split an image into a $5 \times 5$ grid mesh such that it aligns well with the third lines and each of the third lines intersection is right inside one of its cells, as indicated by the red dots.

We collected a set of 2089 images that respect the rule of thirds as the positive set and a set of 2051 images that does not respect the rule of thirds as the negative set. This collection includes images from *Flickr*[2] and *Photo.net*[3]. We randomly allocate 75% of this collection into a training set and the rest 25% into a testing set. For each of the following tests, we repeat the random partition for 10 times and report the average result.

In our method, we use four different saliency maps, namely FT [8], GBVS [6], GC [9], and OBJ (objectness map) [10]. For each saliency map, we extract three types of features, namely the (saliency) centroid, third map, and raw saliency map. We test the effectiveness of each feature using the Naïve Bayesian Classifier. Figure 7(a) shows the precision-recall curves with different features. For each, we combine the same type of feature from all the four saliency maps. We can see that the raw saliency map performs best and the saliency map centroid is least effective, as also shown in the accuracy table (Table I). One reason that the saliency map centroid is not as effective as the other two features is its difficulty to select a suitable window to calculate the centroid, as also discussed in Section II-A. While we try to select an optimal window, our method of picking a fixed threshold to find the window is still not satisfactory. Intuitively, different images shall have different optimal thresholds (windows). As for the thirds map, one reason for its slightly worse performance than the raw saliency map is that the rule of thirds only suggests placing important content along the thirds lines or their intersections, so in practice important content, especially a big object, often extends to the image center from the thirds lines. So dividing an image into a third map that aligns well with the thirds lines is sometimes not as effective as expected.

We also test the effectiveness of each saliency map.

[1]http://opencv.willowgarage.com/wiki
[2]http://www.flickr.com
[3]http://www.photo.net

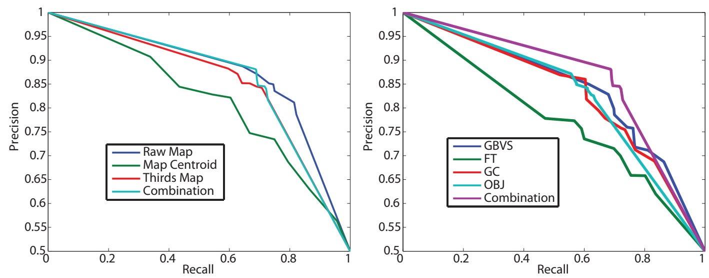


(a) 1st eigen vector (b) 2nd eigen vector (c) 3rd eigen vector (d) 4th eigen vector (e) 5th eigen vector

Figure 6. The most significant eigen-vectors.

| | Naïve Baysian | SVM | kNN | Adaboost |
|---|---|---|---|---|
| centroid | 72.2% | 72.8% | 72.2% | 74.7% |
| thirds map | 77.3% | 76.8% | 75.5% | 79.5% |
| raw saliency map | 78.0% | 79.0% | 78.2% | 80.1% |
| all | 79.1% | 79.9% | 77.7% | 80.5% |

Table I
RULE OF THIRDS DETECTION ACCURACY WITH DIFFERENT FEATURES AND MACHINE LEARNING ALGORITHMS.




(a) curves for features (b) curves for saliency maps

Figure 7. Precision-recall curves that show the performance of different features and saliency maps for the rule of thirds detection.

Figure 7(b) shows the precision-recall curves with different saliency maps. For each, we use all the three type of features. We can see that all these saliency maps basically perform similarly with the FT map a little worse than the others. Combining all these maps gives a slightly better result.

Finally, we show the classification accuracy with each individual feature and their combination using a range of machine learning methods in Table I. Overall, our method achieves around 80% of accuracy in detecting the rule of thirds in a photo.

### *A. Discussion*

Detecting the rule of thirds from a photo is challenging as it requires to identify important content in the photo. Without semantic content understanding, this is necessarily a difficult task like many other multimedia analysis problems. Since our method relies on saliency analysis to infer important content location, our detection method can be mislead by saliency analysis. This contributes to most of the false detection results. Figure 8 shows several failure examples. Our method misclassified the images at row 1 and row 2 as non-rule of thirds images as neither the saliency map nor the objectness map can identify its important content. Our method misclassified the image at row 3 a rule of thirds image as saliency analysis highlighted a part of the object that happens to be around the thirds intersection. For this particular example, the object of interest is at the image center; however, its end is highlighted in the saliency map, which leads our detection method to consider it a rule of thirds image.

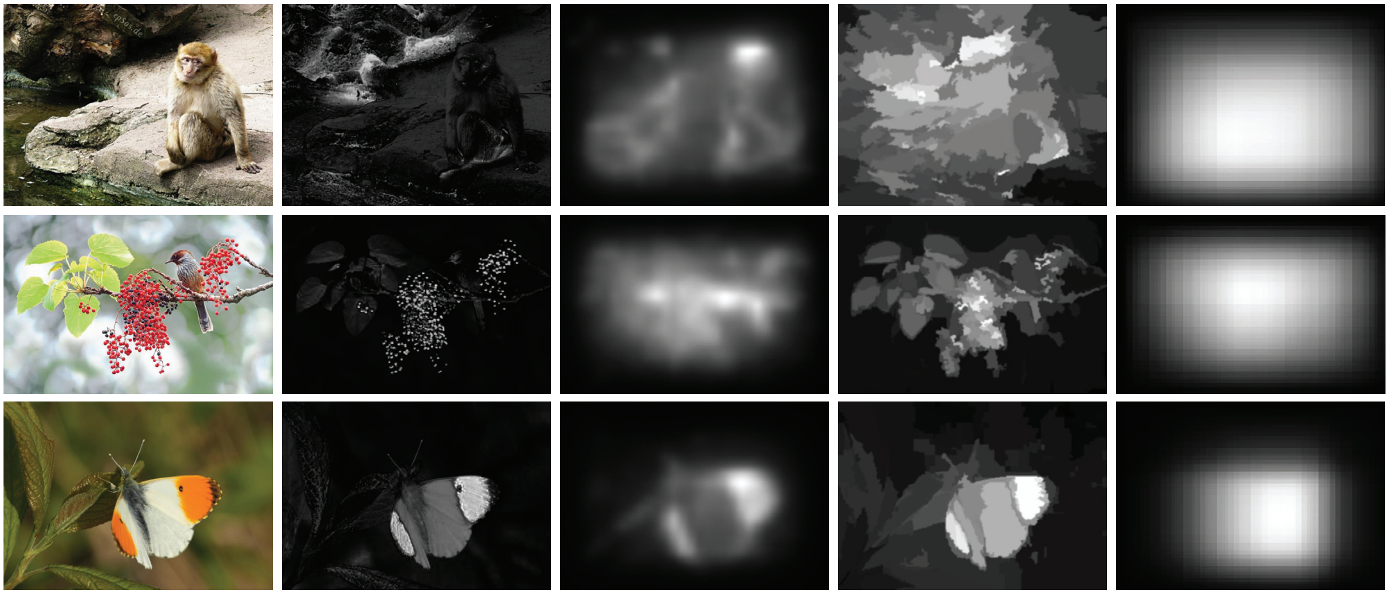
Our method makes use of a generic objectness analysis method as a complement to saliency analysis. However, our experiment showed that the objectness map and the saliency maps are often correlative, so it cannot significantly augment the performance of saliency analysis. Generic object detection is a very challenging task. We believe that the advance in generic object detection will improve our method significantly.

## IV. CONCLUSION

In this paper, we presented an approach for detecting the rule of thirds from a photo. Rule of thirds detection is important for multimedia quality assessment, especially for images and videos. However, this is a difficult task since it requires to automatically identify important content, which is beyond the state of the art in multimedia and computer vision research. This paper examined the state of the art methods in saliency analysis and generic objectness analysis to infer important object locations and designed features accordingly. Our experiments show a promising result although future improvement is certainly needed. In light of this, we plan to finalize our data set and make it public to encourage the research on this new topic.

## ACKNOWLEDGMENT

We would like to thank reviewers for their constructive suggestions. The Flickr images are used under a Creative Commons license from Flickr users: Stuck in Customs, TinkerBells, ∼jjjohn∼, Pyconotus, Storm Crypt, epSos.de, John&Fish, Deanster1983, and steve-h. This research was sponsored in part by the Portland State University Faculty Enhancement Grant.



(a) Input image (b) FT map (c) GBVS map (d) GC map (e) OBJ map

Figure 8. Failure examples. The performance of our method depends on saliency analysis. For images at row 1 and row 2, our method mis-classified them as non-rule of thirds images because saliency analysis cannot identify important objects. Our method misclassified the image at row 3 as a rule of thirds image although the main object is at image center. For this particular example, the left part of the object is highlighted in saliency map instead of the whole object.

## REFERENCES


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

[2] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, no. 2, pp. 469–478, 2010.

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

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

[5] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[6] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of Neural Information Processing Systems*, 2006.

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

[8] R. Achantay, S. Hemamiz, F. Estraday, and S. Susstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1597 – 1604, 2009.

[9] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 409–416.

[10] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object ?" in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 73–80.

[11] F. Liu and M. Gleicher, "Video retargeting: automating pan and scan," in *Proceedings of the international conference on Multimedia*, 2006, pp. 241–250.

[12] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proceedings of ACM SIGGRAPH ASIA 2009 Courses*, 2009, pp. 11:1–11:13.

[13] F. C. Crow, "Summed-area tables for texture mapping," in *Proceedings of ACM SIGGRAPH '84*, 1984, pp. 207–212.

[14] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*. Springer, 1995, pp. 23–37.

[17] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, May 2011.