Compositional Rule of Thirds Detection

Matija Maleš, Adam Heđi, Mislav Grgić
University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
matija.males@gmail.com

***Abstract* -** According to the latest Internet usage statistics, more than 4.5 million photos are being uploaded to Flickr every day. An automated system that can provide feedback about aesthetic value or quality based on learned rules could be a very useful support in picture searching, sorting and editing. This is a challenging problem as it requires semantic understanding of images, which is beyond the state-of-the-art in computer vision. In this paper, we present a saliency-based method for compositional rule of thirds detection. We make use of discriminant analysis to determine whether a photo respects the rule of thirds. The method is faster than previous methods and requires no tuning of many threshold control values and parameters. The results of the experiment demonstrate the validity and efficiency of the proposed method.

Keywords – Aesthetics, Photographic Images; Photography Composition; Rule of Thirds

I. Introduction

Given that many Web-based image repositories are multibillion images in size and that they are rapidly growing every day [1], many image hosting and sharing Web sites have recognized the need for introducing some form of aesthetic or appeal measure [2]. In the last few years, a significant number of articles have been published related to the automatic aesthetic assessment of photographic images [3-8]. Even tools for automatically determining the aesthetic value of photographic images are publicly available [9, 10]. Among many elements that contribute to photograph's aesthetic value are lighting, focus, colors, tonality, content and composition. Of those elements, understanding composition with current computer vision techniques is a highly challenging task.

The rule of thirds is the most common compositional tool, developed by painters' centuries ago and commonly used by professional photographers today. It is one of the most frequently used ways of directing the viewer's eye to the centre of interest in a photographs or paintings. The idea is to imagine the frame split into nine equal sections or rectangles by two horizontal lines and two vertical lines. By placing important compositional elements along these lines or their intersections you will lead viewer's eye through the image and create a more balanced composition. Besides, the image will be more aesthetically pleasing. Two examples are given in Figure 1.

Reliable photographic composition detection is of great importance for measuring aesthetic value of photographic images. In our work we experimented with two methods for salient regions detection and developed features based on saliency maps. Modern classification techniques were used for the rule of thirds detection using these features. Observations and results gained show that our method is robust and comparable to the best existing methods for rule of thirds detection.

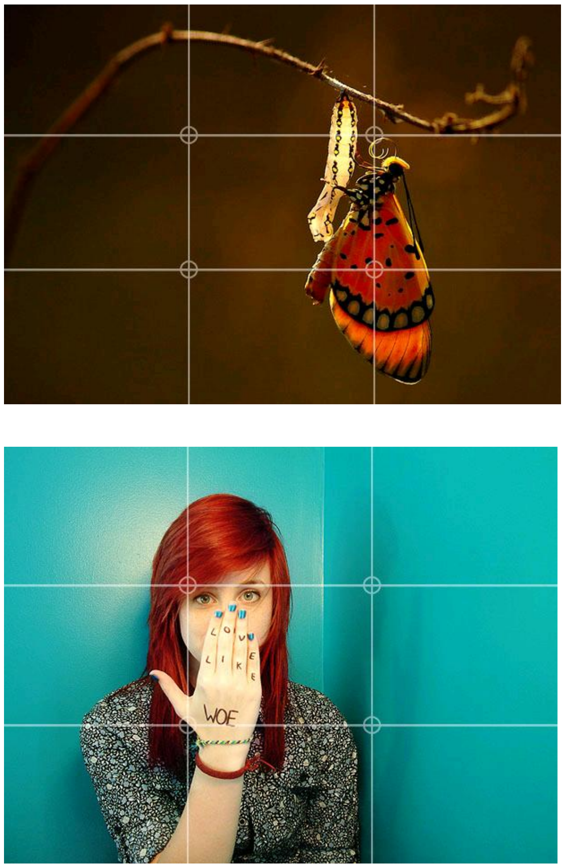

Figure 1. Examples of rule of thirds being used effectively in different types of photographic images.

II. Related Work

Dhar et al. [3] made use of the salient region detector to calculate the minimum distance between the center of mass of the predicted saliency mask and the 4 intersections of third-lines. They calculated the minimum distance to any of the third lines as well. The product of these two numbers (scaled to the range [0, 1]) was used to predict whether an image follows the rule of thirds. Liu et al. [4] proposed similar algorithm, but instead of product they used weighted sum of these two numbers.

Browsing through a large number of professional photographs, Datta et al. [5] observed that a large part of the

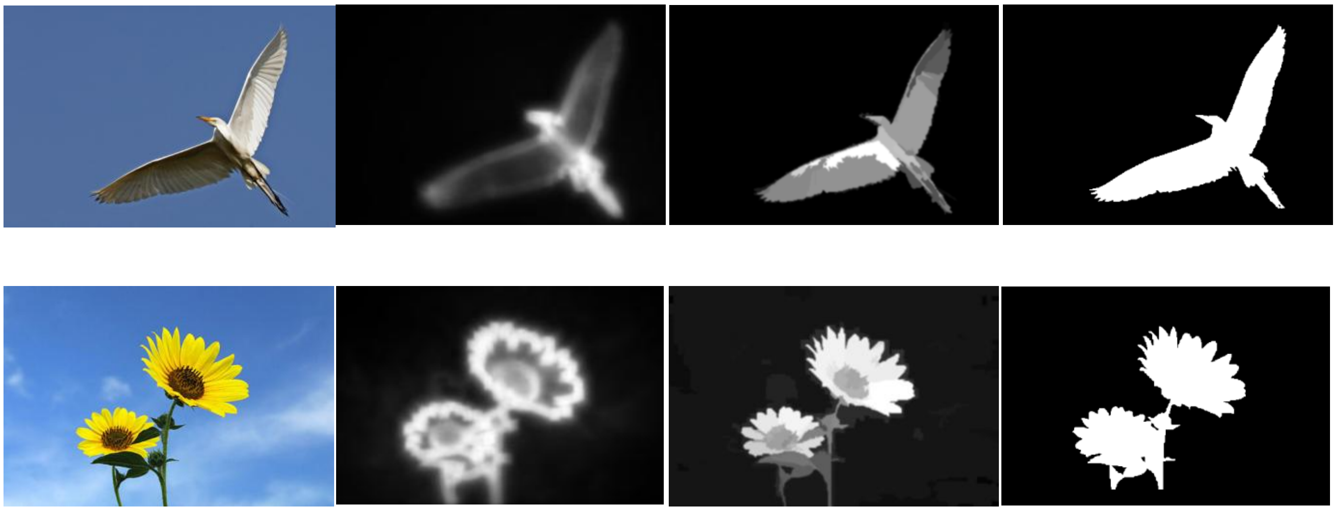

Figure 2. Image saliency maps computed using two different state-of-the-art methods: original images (left), saliency maps computed using CA method (left-middle), saliency maps computed using RC method (right-middle) and saliency cuts (right).

main object often lies on the periphery or inside of the inner rectangle defined by the rule of thirds lines. Based on these observations, they computed the average hue, saturation and value for the inner third, and used these feature vectors to classify image composition.

Our method is most similar to Mai et al. [11], who explored the recent saliency and objectness analysis methods for photography rule of thirds detection. They created a number of features based on the saliency and objectness maps to detect visual elements and their spatial relationship. Well known machine learning algorithms were adopted for the rule of thirds detection using these features. While Mai et al. made use of Naive Bayes Classifier, Support Vector Machine, Adaboost and K-Nearest Neighbor method, in our work we made use of different Discriminant Analysis methods.

III. Salient Region Detection

While humans can detect salient regions from the images fast and reliable, for machines it is a non-trivial task. This is due to the fact that semantic understanding of images is required. In recent years many good and stable salient region detectors have been developed and we made use of them.

In our work we used Context-Aware (CA) salient region detector developed by Goferman et al. [12] which follows four basic principles of human visual attention. In accordance with local low-level considerations, areas that have distinctive colors or patterns should obtain high saliency. Conversely, homogeneous or blurred areas should obtain low saliency values. In agreement with global considerations, frequently occurring features should be suppressed. According to visual organization rules, the salient pixels should be grouped together, and not spread all over the image. Finally, high level principles are implemented as post-processing.

We also used Global Contrast (GC) based salient region detector developed by Cheng et al. [13], specifically its Region Contrast (RC) variant. In RC, input image is first segmented into regions, then color contrast at the region level is computed, and the saliency for each region as the weighted sum of the region's contrasts to all other regions in the image is defined. The weights are set according to the spatial distances with farther regions being assigned smaller weights. RC based saliency cuts (RCC) are obtained by iteratively applying GrabCut [14] to refine the segmentation result, initially obtained by thresholding the saliency map.

Figure 2 shows image saliency maps computed using two different methods. CA saliency maps have higher saliency values at object boundaries, while RC saliency maps have saliency values in salient object regions piece-wise smooth.

IV. Dataset

We downloaded the image dataset collected by Mai et al. [11]. To the best of our knowledge it is the only publicly available collection of images related to the rule of thirds detection problem. This collection includes images from *Flickr.com* and *Photo.net*. However, since some users have removed their images from the before mentioned Web sites, our dataset contains a positive set of 2013 images that respect the rule of thirds and a negative set of 1997 images that do not. Classification of images as being positive or negative was performed by the authors. To test the classifiers on our images, we separated the images into two subsets: the first one is the training set and the second one is the test set. We chose to take at random 75% of the images as a training set, and the rest 25% of the images as a test set. Special care was taken in order to have equal distribution (50%) of images that respect the rule of thirds and those that do not respect the rule of thirds in each set.

V. Rule of Thirds Detection

After applying CA and GC saliency detection algorithms on the input images, saliency maps were created. Saliency cuts

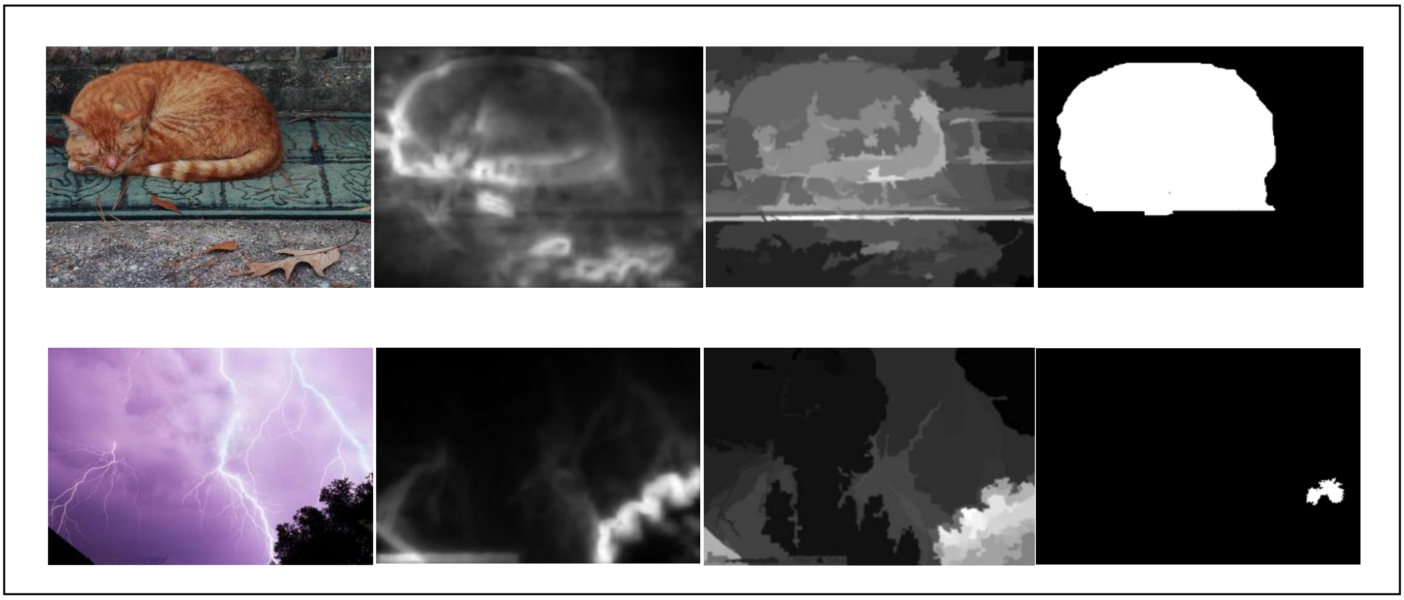

Figure 3. Image saliency maps with misclassified salient regions: image in the first row was classified as a non-rule of thirds image, while image in the second row was classified as a rule of thirds image.

are a special case of GC saliency maps, as described in [13]. Each saliency map was resized into a 20×20 saliency map, and converted into a single feature vector. A matrix containing all maps expressed as feature vectors in its columns was created. Principal Component Analysis (PCA) [15] was applied on the training set to compute the first 15 principle components, which were then used to reduce the feature dimensionality of all the data. The number of principal components used was first determined experimentally via cross validation in [11] and verified in our experiments. These new feature vectors of the training data were used to train classifiers by using popular classification techniques.

The idea was to deduce classification rules from the training set, in order to allow one to classify a new saliency map, i.e. to assign it to one of the classes. There are only two possible classes of images: one class that respects the rule of thirds and the other class that does not. The objective of the classification was to take a decision among these two possibilities. This kind of classification is also called binary classification. Using newly created classification rules we made prediction on the test data.

VI. Discriminant Analysis

Discriminant Analysis (DA) separates samples into classes by finding directions which maximize the variance between classes and minimize the variance within classes. Three discriminant analysis models were compared in our work: Linear Discriminant Analysis (LDA), Mahalanobis Linear Discriminant Analysis (MLDA) and Quadratic Discriminant Analysis (QDA) [16]. LDA, MLDA and QDA are three classic classifiers, with a linear and quadratic decision surface, respectively. LDA (based on the Euclidean distance metrics) and MLDA (based on the Mahalanobis distance metrics) can only learn linear boundaries, while QDA can learn quadratic boundaries and is therefore more flexible. These classifiers are attractive because they have closed form solutions that can be easily computed, are inherently multi-class, and have proven to work well in practice. Also there are no parameters to tune for these algorithms.

We also used Support Vector Machines (SVM) [17], a group of supervised learning models that can be applied for the classification, and finally compared the success rates of the different classifiers. In our work we used the sigma-tuned RBF-kernel model with $\gamma = 3.7$ and cost = 1.0. For DA and SVM models we used Matlab R2011b implementation.

VII. Experimental Results

Effectiveness of the proposed method was evaluated on the before mentioned image dataset collected by Mai et al. [11].

Two state-of-the-art saliency detection algorithms were compared, but related only to the rule of thirds detection. More detailed comparison related to the salient object detection can be found in [13]. In both cases we used the authors' implementations: CA was implemented in Matlab, while RC was implemented in C++. Experiments indicate the RC saliency maps help produce better results, making RC superior to the other algorithm. Also, average time taken to compute RC saliency map was significantly lower compared to the CA saliency map, making it more suitable for a wide range of real-time applications. However, the fact that algorithms are implemented in different programming languages should be taken into account.

Table 1 shows results obtained using different classification methods. It must be noted that the distribution of images in the training and testing set affects the performance of all methods. The final accuracy presented is an average over 10 repeated measurements. In general, LDA and MLDA achieved a lower accuracy than the QDA classifier. The latter method has achieved most significant result with classification accuracy of 77.71 % and can solve the rule of thirds detection problems in a fast and efficient manner. For the purpose of comparison, we compared QDA and SVM. The results indicate that QDA can perform comparable to SVM.

TABLE 1 Results obtained using different classification techniques.

	LDA	MLDA	QDA	SVM
CA	65.72 %	74.02 %	73.98 %	74.07 %
RC	72.67 %	60.88 %	77.71 %	78.61 %
RCC	68.47 %	70.10 %	72.10 %	72.20 %

We found almost no significant differences in classification accuracy between these two classifiers. Without appropriate semantic understanding of image content, rule of thirds detection is a complicated task. Since our method is based on saliency detection algorithms, results are strongly affected by misclassified salient regions. Two misclassified examples are shown in Figure 3.

VIII. CONCLUSION

In this paper we proposed a novel method for compositional rule of thirds detection. This is a challenging problem as it requires semantic understanding of images, which is beyond the state-of-the-art in computer vision. Our method is a saliency-based method which makes use of discriminant analysis to determine whether a photo respects the rule of thirds. The method is faster than previous methods and requires no tuning of many threshold control values and parameters. Extensive experiments on the publicly available dataset demonstrated the effectiveness of the proposed method. The performance of our method in terms of binary classification accuracy is comparable to the best performance to date. We hope that our work will motivate research groups in this new and practically important and challenging research direction.

REFERENCES

[1] royal.pingdom.com/2012/01/17/internet-2011-in-numbers, Pingdom, Internet 2011 in numbers, 2012.

[2] D. Joshi, R. Datta, Q. Luong, E. Fedorovskaya, J. Z. Wang, J. Li and J. Luo, "Aesthetics and Emotions in Images: A Computational Perspective", IEEE Signal Processing Magazine, Vol. 28, No. 5, 2011, pp. 94-115

[3] S. Dhar, V. Ordonez and T. Berg, "High Level Describable Attributes for Predicting Aesthetics and Interestingness", IEEE CVPR 2011, Colorado Springs, USA, 2011, pp. 1657 - 1664

[4] L. Liu, R. Cheny, Lior Wolfz and D. Cohen-Orx, "Optimizing Photo Composition", Computer Graphics Forum, Vol. 29, Issue 2, 2010, pp. 469-478

[5] R. Datta, D. Joshi, J. James and Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach", ECCV 2006, Graz, Austria, 2006, pp. 288-301

[6] Y. Ke, X. Tang and F. Jing, "The Design of High-Level Features for Photo Quality Assessment", IEEE CVPR 2006, New York, NY, 2006, pp. 419–426

[7] C. Li, A. C. Loui and T. Chen, "Towards Aesthetics: A Photo Quality Assessment and Photo Selection System", ACM Multimedia Conference, 2010, pp. 827-830

[8] Y. Luo and X. Tang, "Photo and Video Quality Evaluation: Focusing on the Subject", ECCV 2008, Marseille, 2008, pp. 386-399

[9] R. Datta and J. Z. Wang, "ACQUINE: Aesthetic Quality Inference Engine - Real-time Automatic Rating of Photo Aesthetics", ACM International Conference on Multimedia Information Retrieval, Demonstration, Pennsylvania, USA, 2010, pp. 421-424

[10] L. Yao, P. Suryanarayan, M. Qiao, J. Z. Wang and J. Li, "OSCAR: On-Site Composition and Aesthetics Feedback through Exemplars for Photographers", International Journal of Computer Vision, Vol. 96, No. 3, 2012, pp. 353-383

[11] L. Mai, H. Le, Y. Niu and F. Liu, "Rule of Thirds Detection from Photograph", IEEE ISM 2011, California, USA, 2011, pp. 91 - 96

[12] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-Aware Saliency Detection", IEEE CVPR 2010, California, USA, 2010, pp. 2376 - 2383

[13] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang and S. M. Hu, "Global Contrast based Salient Region Detection", IEEE CVPR 2011, Colorado Springs, USA, 2011, pp. 409 - 416

[14] C. Rother, V. Kolmogorov and A. Blake, "Grabcut - Interactive foreground extraction using iterated graph cuts", ACM Transactions on Graphics, ACM SIGGRAPH, Issue 23, 2004, pp. 309-314

[15] L. I. Smith, "A tutorial on Principal Components Analysis", 2002. www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[16] G. J. McLachlan, "Discriminant Analysis and Statistical Pattern Recognition", Wiley, New Jersey, 2004.

[17] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", 1998. research.microsoft.com/pubs/67119/svmtutorial.pdf