MODELING PHOTO COMPOSITION AND ITS APPLICATION TO PHOTO RE-ARRANGEMENT

Jaesik Park, Joon-Young Lee, Yu-Wing Tai and In So Kweon

Korea Advanced Institute of Science and Technology

ABSTRACT

We introduce a learning based photo composition model and its application on photo re-arrangement. In contrast to previous approaches which evaluate quality of photo composition using the rule of thirds or the golden ratio, we train a normalized saliency map from visually pleasurable photos taken by professional photographers. We use Principal Component Analysis (PCA) to analyze training data and build a Gaussian mixture model (GMM) to describe the photo composition model. Our experimental results show that our approach is reliable and our trained photo composition model can be used to improve photo quality through photo re-arrangement.

Index Terms— Photo composition, Photo re-arrangement

1. INTRODUCTION

Photo composition refers to a set of photography guidelines [1], such as the rule of thirds, the golden ratio, etc, which assists photographers to take professional pleasurable photos. Photo re-arrangement is a set of post-processing techniques for improving photo appearance through cropping and/or re-targeting. Recent representative works of photo composition and photo re-arrangement include Obrador *et al*. [2], Bhattacharya *et al*. [3], Liu *et al*. [4], Judd *et al*. [5] and Cheng *et al*. [6].

In Obrador *et al*. [2], they build an image aesthetic classifier from dominant components of each color segment to measure a visual balance of image features in an image. Bhattacharya *et al*. [3] build a visual composition feature vector using a support vector regression model. Since their method works with user interactions, it recommends an admirable photo composition during re-arrangement. Liu *et al*. [4] utilize the photo composition guidelines and find a cropped and/or re-targetted image frame that maximizes their aesthetic score. Judd *et al*. [5] use machine learning methods to train a bottom, top-down model of visual saliency using multiple image features. Cheng *et al*. [6] use high dimensional features such as color histogram, texture, spatial co-occurrence and prior knowledge of foreground objects to train a classifiers from professional photos for editing an omni-context image.

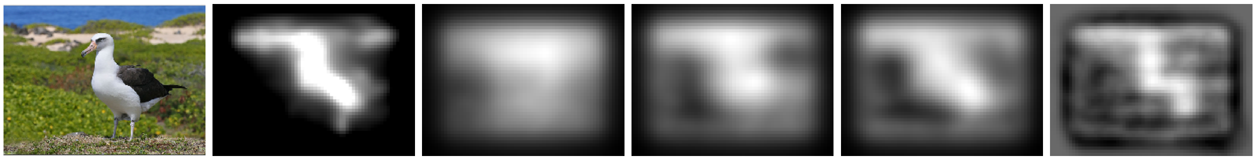
In this work, we introduce a computational method for evaluating photo composition and an application for photo re-arrangement. Our approach is categorized into the top-down approach which models general set of photos. In contrast to the works from Judd *et al*. [5] and Cheng *et al*. [6], we focus on modeling spatial distributions of saliency since we regard it as a key evidence of photo composition. Our method is a data-driven approach that analyzes responses of saliency from a set of pleasurable photos directly. Hence, in contrast to the previous methods [2, 3, 4], our approach does not depend on photo composition guidelines that can be easily biased by a selection of photo composition rules and/or user parameters that adjust the weight balance between different rules. Since our method is data-driven, we can obtain different styles of photo re-arrangement results with different sets of training data.

2. MODELING PHOTO COMPOSITION

We consider an image saliency map is highly correlated to the photo composition guidelines since it represents locations of salient objects in a photo which usually tends to follow human fixations. Our approach utilizes a graph-based saliency detector proposed by Harel *et al*. [7] to get the saliency map from an image. In Harel *et al*.'s method [7], Markov chains were applied to measure similarities between every pair of graph nodes. They define the similarity between adjacent nodes using responses from linear filters. We denote $S(x, y) \in \mathbb{R}^2$ as a saliency map of an image $\mathbf{I}$. Fig. 1 (b) shows an example of S estimated from an image in Fig. 1 (a).

We collect many photos from professional photographers that have good photo compositions to build our photo composition model. Since most digital photos have 4:3 aspect ratio, we normalize the size of saliency maps into a size that have 4:3 aspect ratio. If the aspect ratio of a training image is different from 4:3, we crop the central region of the image to get the 4:3 aspect ratio. In this work, we empirically re-size the saliency map to 64×48 for efficient computation. After that, the saliency map S is vectorized.

We describe the photo composition of the i^{th} image in a training dataset by a feature vector $\mathbf{s}_i$. To produce a compact representation for efficient computing in photo re-arrangement, we stack $\mathbf{s}_i$ and analyze the variation of $\mathbf{s}_i$ using the Principal Component Analysis (PCA). Fig. 2 shows a plot where the first 20 principal components from PCA is able to



(a) (b) (c) (d) (e) (f)

Fig. 1. Reconstruction of saliency map. (a) An image from our training data. (b) Corresponding saliency map. (c – f) Reconstructed saliency maps using 5, 12, 20 and 50 principal components from the learnt photo composition model. It shows an over-fitting result when too many components were used in (f).

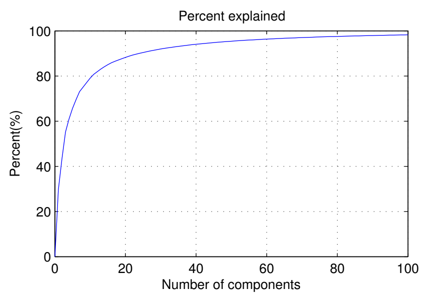


Fig. 2. The plot shows a relation between the number of components and the percent of the variance. We empirically select 20 significant components for saliency map description.

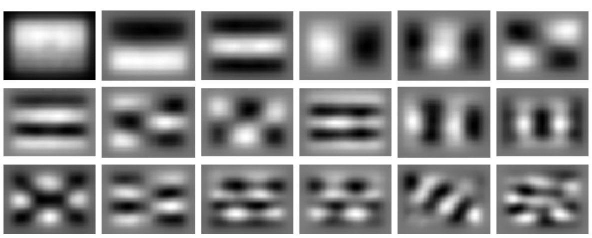


Fig. 3. This figure shows the first 18 principal components of the PCA result computed with one of our dataset. It is normalized for the visualization. The most significant five components are horizontal or vertical components. This shows that significant regions of a common image usually span vertical or horizontal areas of the image.

capture 88.2% of the variation of the training dataset. Fig. 1 shows the reconstructed saliency map using different number of principle components. It shows a good approximation of the original saliency map by only using the first 20 principal components. From these observation, we used the first 20 components with the largest significance values to represent our saliency map. Fig. 3 shows the first 18 components of the PCA result computed with one of our training dataset.

From the projected 20 dimensional training vectors $\mathbf{s}'_i = \mathcal{P}\mathbf{s}_i$, where $\mathcal{P}$ is a projection matrix composed of the first 20 principal components, we fit a Gaussian mixture model $\mathcal{N}$ using an expectation-maximization (EM) algorithm to get our photo composition model.

3. PHOTO RE-ARRANGEMENT

After modeling a photo composition using a GMM $\mathcal{N}$, we can apply $\mathcal{N}$ to photo re-arrangement. The goal of photo re-arrangement is to find a sub-region of original image which the saliency map of the sub-region image is well suited to our photo composition model with a good arrange of salient objects. Compare our approach with the approach from Liu *et al*. [4], we use a statistical model learnt from training data which can handle diverse composition rules implicitly, while Liu *et al*.use a set of predefined measurements from photo composition guidelines which is heuristic and it can be easily biased by user selected parameters.

We parameterize sub-regions of an image plane $\mathbf{I}$ using a sliding window $\mathcal{W}$. The sliding window $\mathcal{W}$ has 4:3 aspect ratio and it is described by a parameter set $\tau = (s, \alpha, \mathbf{t})$ where $s \in [0, 1]$ is a relative scale to the original image, $\alpha \in [-\pi, \pi]$ is a rotation angle and $\mathbf{t} \in \mathbb{R}^2$ is a translation vector. We denote the sub-region $\mathcal{W}(\tau)$ of S as $S_{\mathcal{W}(\tau)}$.

We formulate our solution using a maximum a posteriori (MAP) framework to evaluate τ for the given photo composition model $\mathcal{N}$ and a natural photo composition prior $\mathcal{B}$:

$$\begin{aligned} \tau_{MAP} &= \operatorname*{argmax}_{\tau} P(S_{\mathcal{W}(\tau)}|\mathcal{N}, \mathcal{B}) \\ &= P(\mathcal{N}|S_{\mathcal{W}(\tau)})P(\mathcal{B}|S_{\mathcal{W}(\tau)})P(S_{\mathcal{W}(\tau)}). \end{aligned} \quad (1)$$

The first term $P(\mathcal{N}|S_{\mathcal{W}(\tau)})$ is a likelihood of the saliency vector $\mathbf{s}$ with respect to the GMM $\mathcal{N}$ that is determined in Sec. 2. The likelihood is defined as

$$P(\mathcal{N}|S_{\mathcal{W}(\tau)}) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathcal{P}\mathbf{s}_{\mathcal{W}(\tau)}, \mu_k, \Sigma_k), \quad (2)$$

where $\mathbf{s}_{\mathcal{W}(\tau)}$ is a normalized saliency vector of $S_{\mathcal{W}(\tau)}$.

The second term $P(\mathcal{B}|S_{\mathcal{W}(\tau)})$ is a prior of a natural photo composition. In the previous work by Judd *et al*. [5], they analyzed a large scale eye-tracking dataset and found out humans tend to gaze at the central region of an image. This observation introduced the central region prior for the saliency detection. Inspired by Judd*et al*.'s work [5], we define our prior function C as condensation of saliency magnitudes in central region. In addition to this prior, we introduce a global prior function G as a relative amount of saliency in the given sub-region of an image to the whole magnitude of the image. This

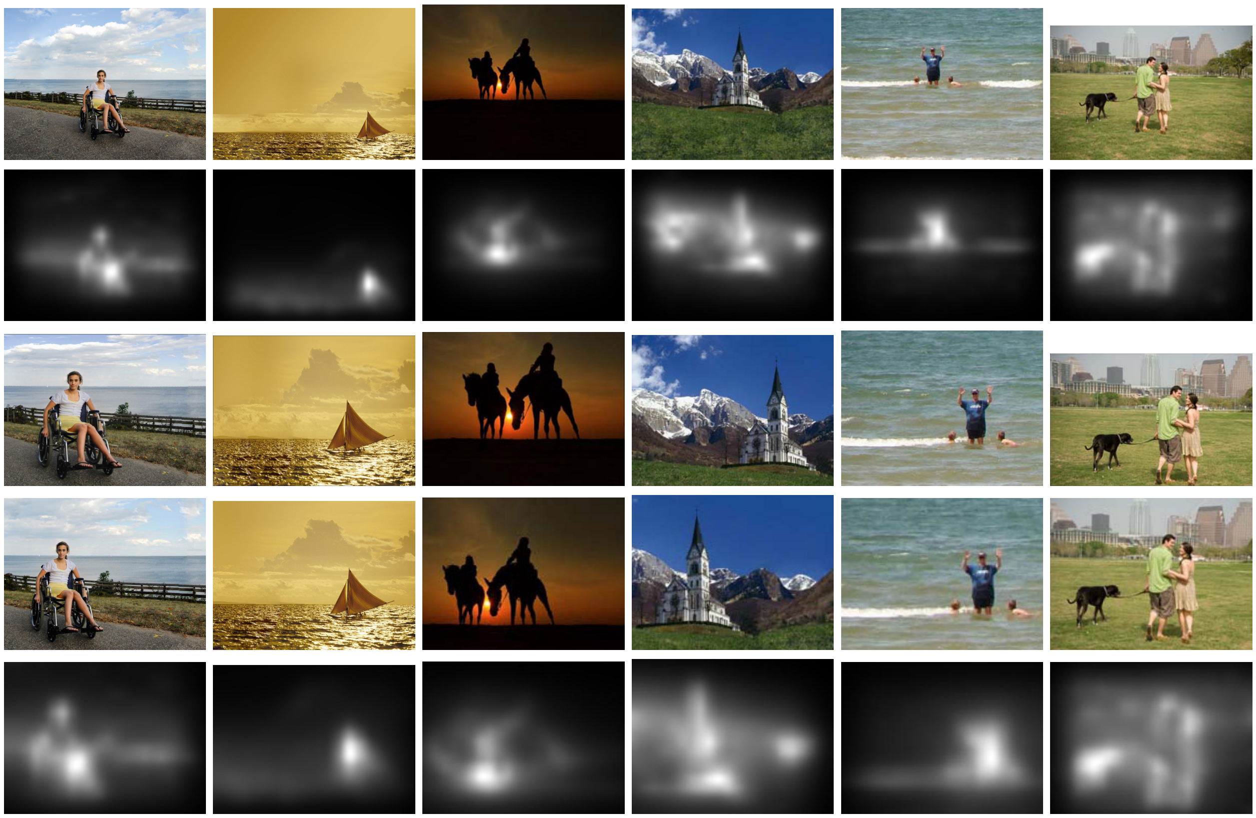


Fig. 4. Experimental results on image re-arrangement. First row: Input images. Second row: Corresponding saliency map. Third row: Results from Liu *et al*. [4]. Fourth row: Results using our approach. Fifth row: Saliency map of re-arranged photo

additional term prevents a bias that focuses a specific salient region while loosing the context of an image. Thus, our prior term is defined as

$$P(\mathcal{B}|S(s,\alpha,\mathbf{t})) = \frac{C(S_{\mathcal{W}(\tau)})G(S_{\mathcal{W}(\tau)})}{M}, \tag{3}$$

where M is a normalization factor, and the function C and G are

$$C(S_{\mathcal{W}(\tau)}) = \frac{\sum_{x,y\in\mathcal{W}_c(\tau)} S(x,y)}{\sum_{x,y\in\mathcal{W}(\tau)} S(x,y)}, \tag{4}$$

$$G(S_{\mathcal{W}(\tau)}) = \frac{\sum_{x,y\in\mathcal{W}(\tau)} S(x,y)}{\sum_{x,y\in\mathbb{I}} S(x,y)}. \tag{5}$$

$\mathcal{W}_c(\tau)$ is the central region of $\mathcal{W}(\tau)$. In our implementation, we set $\mathcal{W}_c(\tau)$ as a rectangular region which is smaller than the $\mathcal{W}(\tau)$ by a factor of 0.8. We set the probability $P(S_{\mathcal{W}(\tau)})$ in Eq. (1) to a constant since we assume each specific saliency map have the equal possibility for any parameter set.

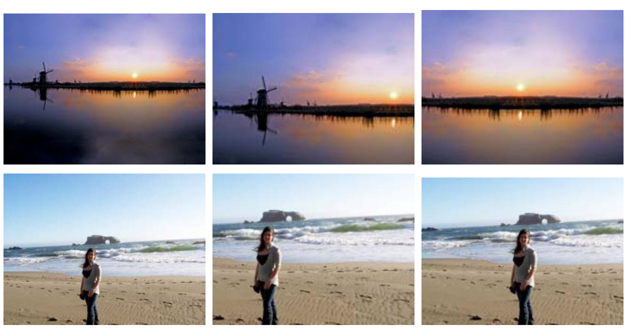
We find a maximum value of Eq. (1) by exhaustive searching in the quantized space of τ. When the maximum value of the posterior $P(S_{\mathcal{W}(\tau)}|\mathcal{N},\mathcal{B})$ is smaller than a certain threshold, we regard that the photo composition of the given image is hard to determine and set the similarity transformation parameters to a default one, $\tau = (1, 0, [0, 0]^T)$.

4. EXPERIMENTAL RESULT

In this section, we present our results on photo re-arrangement. We set the sub-region parameter $\tau = (s, \alpha, \mathbf{t})$, $s \in [0.6, 1]$ and $\mathbf{t} = [\pm 10p, \pm 10q]^T$ as the search space of the optimal sub-region where p and q are arbitrary integer numbers of the pixel unit. For simplicity, we consider only the s and $\mathbf{t}$ in our experiment, but our approach can be easily extended to include rotations into the search space. Our results were obtained using the same parameters setting for all experiments.

Our first experiment uses scenery photos to train the photo composition model. Our training set consists of 3,695 photos which are acquired by using a keyword 'landscape' in `Flickr.com`. We reject images with low popularity since we believe popular photos usually have better aesthetics as well as better photo composition. Fig. 5 shows a subset of images in the 'landscape' dataset. We compare our results with results from Liu *et al.*'s method [4] in Fig. 4. Our results are pleasurable and are similar to the results from Liu *et al.*'s method [4]. Note that we do not model any photo composition guidelines [1] explicitly unlike Liu *et al.*'s method [4]. We believe that the similar photo re-arrangement results are due to the fact that the photo composition of the 'landscape' category usually have a high fidelity of photo composition guidelines such as the rule of thirds, the golden ratio, the golden trian-

Fig. 5. Two subsets of training images what we used. first row: 'landscape' dataset. second row: 'stock photo' dataset.



(a) (b) (c)

Fig. 6. Photo re-arrangement with different training datasets. (a) Input photos. (b) Our results using the 'landscape' dataset. (c) Our result using the 'stock photo' dataset. According to a training set, places of foreground objects are affected.

gles, etc. Our approach can successfully learn these guidelines through our data-driven training process.

Our approach is data-driven, hence, we can perform photo re-arrangement for another category of images using the same framework but with different training set. We collect another 3,415 high quality stock photos ('stock photo' dataset) which contain a main foreground object from various categories. Fig. 5 shows a subset of the 'stock photo' dataset. We use the same test images used in [4] for comparison. The photo composition model using the 'stock photo' dataset is learnt using the same method with the same parameters as the 'landscape' dataset. The photo re-arrangement results with different training sets are shown in Fig. 6. The results show the property of our data-driven approach which can be applied using different training set for different preferences of photo arrangement.

5. CONCLUSION AND FUTURE WORK

In this work, we have introduced a framework to model photo composition and its application to photo re-arrangement for better aesthetics. We verified our method using both the public and our dataset. Our results were compared to the results from the recent work that use the photo composition rules explicitly. Our future work is to develop a general photo re-arrangement system that can convert an arbitrary image into a specific photographic style.

6. ACKNOWLEDGEMENT

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2012-C7000-1001-0007) and the National Research Foundation of Korea (No. 2011-0013349).

7. REFERENCES

[1] P. Jonas, "Photographic composition simplified," *Amphoto Publishers*, 1976.

[2] Pere Obrador, Ludwig Schmidt-Hackenberg, and Nuria Oliver, "The role of image composition in image aesthetics," in *17th IEEE International Conference on Image Processing (ICIP)*, 2010.

[3] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah, "A coherent framework for photo-quality assessment and enhancement based on visual aesthetics," in *ACM Multimedia International conference*, 2010.

[4] Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or, "Optimizing photo composition," *Computer Graphic Forum (Proceedings of Eurographics)*, vol. 29, no. 2, pp. 469–478, 2010.

[5] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba, "Learning to predict where humans look," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[6] Bin Cheng, Bingbing Ni, Shuicheng Yan, and Qi Tian, "Learning to photograph," in *ACM Multimedia International conference*, 2010.

[7] Jonathan Harel, Christof Koch, and Pietro Perona, "Graph-based visual saliency," in *Twentieth Annual Conference on Neural Information Processing Systems (NIPS)*, 2006.