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 \( I \). 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 \( 64 \times 48 \) 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 \( s_i \). To produce a compact representation for efficient computing in photo re-arrangement, we stack \( s_i \) and analyze the variation of \( s_i \) using the Principal Component Analysis (PCA). Fig. 2 shows a plot where the first 20 principal components from PCA is able to
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 $I$ using a sliding window $\mathcal{W}$. The sliding window $\mathcal{W}$ has 4:3 aspect ratio and it is described by a parameter set $\tau = (s, \alpha, t)$ where $s \in [0, 1]$ is a relative scale to the original image, $\alpha \in [-\pi, \pi]$ is a rotation angle and $t \in \mathbb{R}^2$ 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 $\tau$ for the given photo composition model $\mathcal{N}$ and a natural photo composition prior $\mathcal{B}$:

$$\tau_{MAP} = \arg\max_{\tau} P(S_{\mathcal{W}(\tau)}|\mathcal{N}, \mathcal{B})$$

$$= P(\mathcal{N}|S_{\mathcal{W}(\tau)}) P(\mathcal{B}|S_{\mathcal{W}(\tau)}) P(S_{\mathcal{W}(\tau)}). \quad (1)$$

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

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

where $s_{\mathcal{W}(\tau)}$ 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
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

$$P(B|S(\tau)) = \frac{C(S_W(\tau))G(S_W(\tau))}{M}, \quad (3)$$

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

$$C(S_W(\tau)) = \frac{\sum_{x,y \in W_c(\tau)} S(x,y)}{\sum_{x,y \in W(\tau)} S(x,y)}, \quad (4)$$

$$G(S_W(\tau)) = \frac{\sum_{x,y \in W(\tau)} S(x,y)}{\sum_{x,y \in I} S(x,y)}. \quad (5)$$

$W_c(\tau)$ is the central region of $W(\tau)$. In our implementation, we set $W_c(\tau)$ as a rectangular region which is smaller than the $W(\tau)$ by a factor of 0.8. We set the probability $P(S_W(\tau))$ 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 $\tau$. When the maximum value of the posterior $P(S_W(\tau)|N, B)$ 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 $\tau = (s, \alpha, t)$, $s \in [0.6, 1]$ and $t = [\pm 10p, \pm 10q]^T$ 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 $t$ 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-
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