InspireMePosing: Learn Pose and Composition from Portrait Examples

Bin Sheng¹, Yuxi Jin², Ping Li², Wenxiao Wang³, Hongbo Fu³, and Enhua Wu⁴,⁵

¹Shanghai Jiao Tong University, Department of Computer Science and Engineering, China
²Macau University of Science and Technology, Faculty of Information Technology, Macau
³City University of Hong Kong, School of Creative Media, Hong Kong
⁴University of Macau, Faculty of Science and Technology, Macau
⁵Chinese Academy of Sciences, State Key Laboratory of Computer Science, Institute of Software, China

Abstract
Since people tend to build relationship with others by personal photography, capturing high quality photographs on mobile device has become a strong demand. We propose a portrait photography guidance system to guide user’s photographing. We consider current scene image as our input and find professional photograph examples with similar aesthetic features for it. Deep residual network is introduced to gather scene classification information and represent common photograph rules by features, and random forest is adopted to establishing mapping relations between extracted features and examples. Besides, we implement our guidance system on a camera application and evaluate it by user study.

CCS Concepts
● Computing methodologies → Image manipulation; Image processing; ● Applied computing → Arts and humanities;

1. Introduction
Recently, the developing camera technology makes it easier to capture high quality images by mobile devices. However, taking an excellent photograph requires rich experience and enough patience which are what most users lack. It is tough for common users to decide the composition and human posture like experienced photographers. Although existing works can improve the quality of photos took by average person, they cannot make connection among aesthetics features of those scene and give suggestions.

Visual balance is key to achieve good painting and contains composition, color and tone [DJLW06]. Some works segment an image into regions, extract salient regions and find out relative distance and distributions of visual elements by identifying the importance of each regions [LCWCO10, STR⁺04, LWZ⁺17]. And some other works remodel the scene by using content-aware approach [AS12, LWT13]. Bin Cheng et al. [CNYT10] design a method to learn object spatial correlation distribution and use this method to guide the composition of professional photos. Hongbo Fu et al. [FHP13] enrich variety of posing in portrait photography through data-driven suggestions. Yin et al. [YMCL13] develop a socialized mobile photography system, and assist mobile user in capturing high quality photos. Ni et al. [NXC⁺13] learn from a large image data set of professional landscape photos and build an omni-range spatial context model. Recently, using deep neural network to automatic learn feature and train classifier for images has shown promising result [SVZ13, HZRS15, SWT14]. [LLJ⁺15, CCOQ15] investigate automatic feature learning to predict image aesthetics. However, these methods based on networks cannot clearly explain what they have learned from networks, and they cannot visualize features they have learned.

In this paper, we assume that the model posture and photo composition chose by professional photographer are positive on the aesthetics of the scene, and common users should learn how professional photographers make use of model posture and photo composition to get a high quality image. Thus, we design a portrait photography guidance method, which takes current scene image as input, finds professional photograph examples that have similar aesthetic features with the current scene and outputs compositions and postures of professional photograph examples. With the guidance of those professional compositions and postures, users can well improve their capturing results.

2. Our Approach
In this paper, we propose to pre-guide the composition and posture for getting a high quality photo. We help users decide the composition and posture before they take shot. We first extract aesthetic features which are considered as reference during photoing. Then, we train a suggestion model to present guidance for users.

We summarize features of an image by widely accepted photograph rules, and segment an image into regions for indicating the
visual arrangement of objects like [NXC13]. We also present type of scenes as features by a deep learning method [HZRS15]. We propose to estimate illumination condition by studying weak cues from images. We measure the saliency of each pixel belonging to the background of an image by a spectral residual approach saliency detection [HZO7], and apply the pattern for the rule of third to the background of an image by a spectral residual approach saliency detection [HZRS15]. We propose to estimate illumination condition by studying weak cues of scenes as features by a deep learning method [HZRS15]. We conduct two user studies to evaluate the effectiveness of our technique. First, we evaluate whether our model can provide reasonable examples for the input scene image by performing experiments on suggestion model. Then we verify if our method can really help users improve their photograph results by conducting experiments on usage results.

We invite totally 100 students to take part in our study, of which 50 are women and 50 are men. We provide ten sets of scene images for the input scene image from the best example photograph to generate 3 photographs separately. And the total nine photographs are sorted in random order at first.

3. Implement and Result

We conduct two user studies to evaluate the effectiveness of our technique. First, we evaluate whether our model can provide reasonable examples for the input scene image by performing experiments on suggestion model. Then we verify if our method can really help users improve their photograph results by conducting experiments on usage results.

We invite totally 100 students to take part in our study, of which 50 are women and 50 are men. We provide ten sets of scene images and corresponding photographs for each participant during the user study. Participants are asked to sort out the reference photographs for the input scene image from the best example photograph to the worst example photograph. All the three methods in comparison generate 3 photographs separately. And the total nine photographs are sorted in random order at first.

The statistics of the core information was showed by Tab.1. The
photographs provided by our method rank 3.44 on average, which is significantly higher than other method. And the ranking variance is 2.76, which shows that our method performs more stable than other method.

![Figure 2: Average aesthetic degree and match degree. sug denotes model of suggestion, random denotes model of random suggestion and nosug denotes model of no suggestion.](image)

We also design three models for comparison to verify whether our technique is useful or not. Model of suggestion can provide suggestions with our technique. Model of random randomly chooses suggestion. Model of no suggestion provides no suggestion. We conduct the study in two different scene: a park and next to a road. The park scene is at the green land of a park, and the road scene is next to a road.

Users are asked to score the aesthetic degree of the photo and the match degree of the composition, posture of model in the scene with 0 to 100 degree. Fig. 2 shows the average aesthetic degree and match degree of the composition, posture of model for the two different scene: park and road. Model of suggestion reaches the highest average score, and the effect of model of suggestion is better than model of no suggestion.

4. Conclusion

We propose a data-driven method to guide portrait photography. In our method, current scene image is treated as input, and professional photograph examples which have similar aesthetic features as the input scene image are found. The effectiveness and practical of our proposed method have been demonstrated by experiments. We are the first attempting to help users improve their photo’s quality during the photoing procedure. It is possible to improve the effectiveness of our application through recognising objects on background to provide more interactive pose.

References


