Pacific Graphics (2018)
H. Fu, A. Ghosh, and J. Kopf (Editors)

Short Paper

InspireMePosing: Learn Pose and Composition from Portrait Examples

Bin Sheng[1], Yuxi Jin[2], Ping Li[2], Wenxiao Wang[2], Hongbo Fu[3], and Enhua Wu[4,5]

[1]Shanghai Jiao Tong University, Department of Computer Science and Engineering, China
[2]Macau University of Science and Technology, Faculty of Information Technology, Macau
[3]City University of Hong Kong, School of Creative Media, Hong Kong
[4]University of Macau, Faculty of Science and Technology, Macau
[5]Chinese 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, CCQ15] 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

DOI: 10.2312/pg.20181274

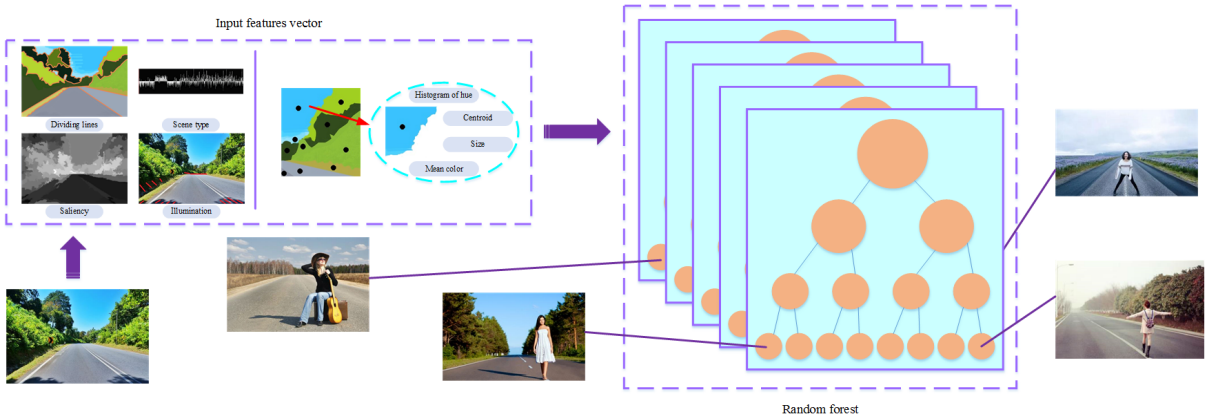

Figure 1: *Prediction for a feature vector in a decision tree. Each leaf in is a photograph shot by a profession photographer and the training target is the reference photo. The aesthetics feature is routed through each node in the tree according to the function ϕ of the node.*

visual arrangement of objects like [NXC*13]. 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 [HZ07], and apply the pattern for the rule of third to the saliency image.

We adopt a graph based segmentation method [FH04] to divide background region into different regions. Both input image and reference image are divided into numbers of regions. We merge regions by a match score evaluating similarity of two regions with extracted features, which can be formulated as following.

$$E(R_1,R_2) = Ke^{\frac{-D_d(R_1,R_2)^2 - D_{SV}(R_1,R_2)^2}{2\times\sigma}} \quad (1)$$

where R_1 and R_2 are two different regions, $\sigma = 0.2$, and

$$K = D_{size}(R_1,R_2) \times histoSim(R_1,R_2)$$

$D_d(R_1,R_2)$ is defined as the arrangement distance between R_1 and R_2.

$$D_d(R_1,R_2) = \frac{d(C(R_1),C(R_2))}{(\sqrt{2}\times 128)} \quad (2)$$

where $d((x_1,y_1),(x_2,y_2)) = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$, $C(R_1)$ and $C(R_2)$ are the centroid of regions R_1 and R_2.

And $D_{SV}(R_1,R_2)$ denotes the SV difference in mean color HSV between R_1 and R_2.

$$D_{SV}(R_1,R_2) = \frac{d((HSV(R_1).s,HSV(R_1).v),(HSV(R_2).s,HSV(R_2).v))}{(\sqrt{2}\times 255)} \quad (3)$$

$D_{size}(R_1,R_2)$ is the size difference between R_1 and R_2, and can be calculated as:

$$D_{size}(R_1,R_2) = \frac{|Size(R_1) - Size(R_2)|}{Size(R_1) + Size(R_2)} \quad (4)$$

The similarity $histoSim(R_1,R_2)$ of the histogram is calculated by adding overlap of each bin from two regions and divided by the sum with the size of larger region. The match regions from reference image for the input image are those with highest match score.

Besides, we use randomized hough transform [XOK90] to extract dominant lines of an image. The dominant lines of the input image and reference image are also matched in our paper.

By recursively branching, a decision tree classifies a feature vector into two branch until reaching a leaf node. Fig.1 shows the prediction for a feature vector in random decision forest. We save the top 5 predictions rather than combine them into one, because photographing is not just a regression problem, and there are kinds of ways to get a good photograph. The final recommendation score S of the reference image, which is used to rank the rest of the reference photographs, is formulated as following.

$$S = Rscore \times (1 + \lambda \times MatchLine) \quad (5)$$

where $Rscore$ is region matching score and $MatchLine$ is the number of matched line from reference images for the input image.

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 method 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

Table 1: *Ranking summary of evaluation*

Groups	Count	Sum	Average	Variance
Our method	2610	8978	3.44	2.76
Random(all candidates)	2610	18192	6.97	4.12
Similar geography	2610	11980	4.59	3.13

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.

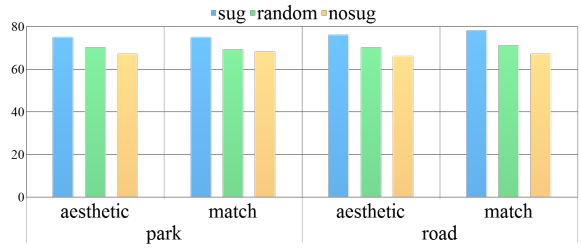

Figure 2: *Average aesthetic degree and match degree. sug denotes model of suggestion, random denotes model of random suggestion and nosug denotes model of no suggestion.*

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 recognisiing objects on background to provide more interactive pose.

References

[AS12] AVIDAN S., SHAMIR A.: Seam carving for content-aware image resizing. *Siggraph 26*, 3 (2012), 10. 1

[CCQ15] CAMPBELL A., CIESIELKSI V., QIN A. K.: *Feature Discovery by Deep Learning for Aesthetic Analysis of Evolved Abstract Images*. Springer International Publishing, 2015. 1

[CNYT10] CHENG B., NI B., YAN S., TIAN Q.: Learning to photograph. In *International Conference on Multimedia* (2010), pp. 291–300. 1

[DJLW06] DATTA R., JOSHI D., LI J., WANG J. Z.: Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision* (2006), pp. 288–301. 1

[FH04] FELZENSZWALB P. F., HUTTENLOCHER D. P.: Efficient graph-based image segmentation. *International Journal of Computer Vision 59*, 2 (2004), 167–181. 2

[FHP13] FU H., HAN X., PHAN Q. H.: Data-driven suggestions for portrait posing. In *SIGGRAPH Asia 2013 Technical Briefs* (2013), pp. 1–4. 1

[HZ07] HOU X., ZHANG L.: Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* (2007), pp. 1–8. 2

[HZRS15] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. 770–778. 1, 2

[LCWCO10] LIU L., CHEN R., WOLF L., COHEN-OR D.: Optimizing photo composition. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 469–478. 1

[LLJ*15] LU X., LIN Z., JIN H., YANG J., WANG J. Z.: Rapid: Rating pictorial aesthetics using deep learning. *IEEE Transactions on Multimedia 17*, 11 (2015), 2021–2034. 1

[LWT13] LUO W., WANG X., TANG X.: Content-based photo quality assessment. *IEEE Transactions on Multimedia 15*, 8 (2013), 1930–1943. 1

[LWZ*17] LIANG Y., WANG X., ZHANG S.-H., HU S.-M., LIU S.: Photorecomposer: Interactive photo recomposition by cropping. *IEEE Transactions on Visualization and Computer Graphics* (2017). 1

[NXC*13] NI B., XU M., CHENG B., WANG M., YAN S., TIAN Q.: Learning to photograph: A compositional perspective. *IEEE Transactions on Multimedia 15*, 5 (2013), 1138–1151. 1, 2

[STR*04] SETLUR V., TAKAGI S., RASKAR R., GLEICHER M., GOOCH B.: Automatic image retargeting. In *Mum'05 : Proceedings of the International Conference on Mobile and Ubiquitous Multimedia, New York, Ny, Usa* (2004), p. 4. 1

[SVZ13] SIMONYAN K., VEDALDI A., ZISSERMAN A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. *Computer Science* (2013). 1

[SWT14] SUN Y., WANG X., TANG X.: Deep learning face representation by joint identification-verification. 1988–1996. 1

[XOK90] XU L., OJA E., KULTANEN P.: A new curve detection method: Randomized hough transform (rht). *Pattern Recognition Letters 11*, 5 (1990), 331–338. 2

[YMCL13] YIN W., MEI T., CHEN C. W., LI S.: Socialized mobile photography: Learning to photograph with social context via mobile devices. *IEEE Transactions on Multimedia 16*, 1 (2013), 184–200. 1