FINDING YOUR SPOT: A PHOTOGRAPHY SUGGESTION SYSTEM FOR PLACING HUMAN IN THE SCENE

Shuang Ma[†‡], *Yangyu Fan*[†], *Chang Wen Chen*[‡]

[†] Northwestern Polytechnical Universty, Xi'an, China
[‡] State University of New York at Buffalo, Buffalo, NY, USA

ABSTRACT

Capturing a professional like photo is always a challenging task, especially for novice users. This paper proposes a photography suggestion approach to assist users to take high visual quality photos with human in the scene. In this research, we first investigate a set of aesthetic composition rules and visual perception principles to construct an aesthetic score prediction model in order to measure visual quality in terms of photo composition. Then we conduct a study on professional photos to define a proper size for the enclosure of human into the picture of a given scene. The proposed approach is able to leverage saliency and geometric detection to represent composition features. Finally, we utilize an efficient hierarchical search to obtain the optimal enclosure for human in the scene. Extensive experiments have been performed for hot spot landmark locations. Through subjective evaluation, the results show that the proposed approach can effectively provide appealing composition recommendation to help users take high quality photos with human in the scene.

Index Terms— Photography suggestion, aesthetic principles, view specific

1. INTRODUCTION

The recent popularity of mobile devices equipped with cameras has revolutionized people's daily life. People are more enthusiastic than before in taking photos to share their experiences and spontaneous moments. However, it usually requires years of practice and experience to take professional grade photos. Recently, a photography suggestion system has been developed to assist smartphone users to take professional grade photos by learning photography rules and through social context [1, 2]. However, the proposed system is only applicable to landscape without human in the scene. Therefore, it is very much desired to develop a photography suggestion system that can tell amateur users to capture high quality photos of tourist landscape with human in the scene.

To accomplish the goal of suggestion, it is necessary to investigate computational aesthetics of photography and photo quality assessment. Existing works have been developed aiming at automatically assessing or enhancing image quality with computational models based on various visual features. Some of them focused on low level image features, such as linear regression of low level features to predict aesthetics rating [3], approximation of photography composition guidelines [4], and an interactive application using spatial recomposition [5]. Others make use of high level knowledge, such as describable image attributes that human might use to evaluate an image [6], resolution and physical dimensions of an image in affecting viewers' appreciation [7], mining the underlying knowledge of professional photographers from massively crawled photos [8], and assessing photo quality with geo-context and crowdsourcing [9].

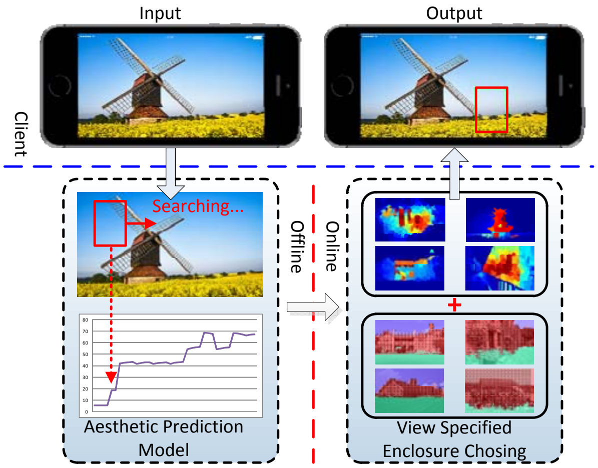
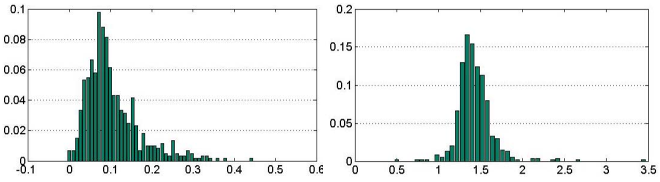


Fig. 1: Overview of the photography suggestion system

However, these existing attempts predominantly focused on how to find a good view enclosure and are mostly restricted to landscape photos. The proposed system is different from previous works in that this system attempts to suggest how to place human in the scene when taking photos of landscape together with human. We call this type of photos landscape-portrait photos. We require the final photo to be able to highlight both human and scene, rather than either normal portrait photos focusing on human or landscape only photos focusing on scenery. Moreover, instead of applying simple rules heuristically, a view specific approach is developed to recommend to the mobile device users at which location should the human be placed in the scenic views mimicking the reasoning of a professional photographer.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE



(a) (b)

Fig. 2: Statistical results of study on professional photographs

2. AESTHETIC PRINCIPLES MODELING

The pipeline of the proposed approach is shown in Fig 1. An aesthetic composition model is constructed offline to generate prediction score for each enclosure. When input a landscape photo, salient and geometric detections are adopted to select candidate enclosures in a specific view. Finally a hierarchical search is carried out to efficiently search an optimal enclosure to guide users in taking professional like photos with human in the scene.

2.1. Problem formulation

Given a landscape image I, the object of the proposed system is to generate a suggestion to guide the users with suitable size $s(w,h)^T$ and position $c(x,y)^T$ for them to place human in the scene with following considerations:

- The suggestion should follow key aesthetic rules to make it appealing and attractive so as to enhance the overall quality.
- Proper size should be defined to be consistent with professional photographers' visual perception.
- Visual perception principles should be applied to place human in a proper region within the photo.
- The suggested position should not be placed outside the border of image I
- The suggested position should not block any salient objects in order to maintain perceptual balance between scenic objects and human in the scene.

Based on the considerations, we formulate the problem of photography suggestion to place human in the scenic view as a search problem in a joint-state-space. We utilize a sliding window as the enclosure finder, and its configuration is described by the two-dimension coordinates of the window center $c = [x, y]^T$, and the size of window $s = [w, h]^T$. Given an input scenic image I, we aim to find an enclosure $R(I; c, s)$ within I, and label it according to the configuration of the enclosure finder defined by c and s. Thus, the problem can be formulated as:

$$\{c^*, s^*\} = \arg\max_{\substack{g_i\in\Phi\\ c,s}} S_{pre}(R(I;c,s)) \cdot \mathbb{1}(c = g_i) \cdot \mathbb{1}(V_s(R(I;c,s)) \leq \gamma) \quad (1)$$

where $\mathrm{S}_{pre}(\cdot)$ is an aesthetic score prediction measure, $\mathbb{1}(\cdot)$ is a discriminant function, whose value is 1 when the condition in side bracket is satisfied, otherwise it is 0. $V_s(\cdot)$ is the saliency value occupied by $R(I; c, s)$, γ is a threshold defined by the constraint that human shall not block the salient object in I, Φ is the detected ground region, which will be explained in Section 3.1, g_i is the i_{th} sliding window' center in Φ.

2.2. Aesthetic prediction score modeling

2.2.1. Defining size based on study of professional photos

To better estimate professional photographer's insight to define the size of human in the scene, we collect a database of professional landscape-portrait photos. This database does not include pure portrait photos that contain no landscape scenes. The professional photos obtained via crawled-sourcing are all collected carefully to contain only landscape-portrait photos. We label each photo by hand in the dataset with bounding box over the human subject(s). Fig 2 shows the histogram of the size of human region (Fig 2a) and height-width ratio (Fig 2b) obtained from the dataset of more than 1000 professional images. We can see from these histograms that the size for human subjects is very well distributed with clear mean values. The parameters of these distributions shall be used to guide the determination of human size and height-width ratio, i.e., initial size s will be 0.1% of I, and the height-width ratio will be set as 3:2.

2.2.2. Aesthetic and visual perception representation and modeling

Composition is an important aesthetic aspect of a photo. There are various aesthetic rules for capturing well-composed photos. We consider such rules to be prominent in most aesthetically appealing photos. To apply these rules computationally, we define a prediction score to evaluate the aesthetics of the suggested composition for taking landscape-portrait photos.

Rule of Thirds

The rule of thirds is one of the most important composition rules used by professional photographers to capture appealing aesthetic photos. It encourages placement of important objects along the imagery thirds lines or around their intersections. The rule of thirds prediction score S_{RT} is calculated as:

$$S_{\mathrm{RT}}=\omega_{vi}100 \times [exp(-\frac{D_{RT}}{diag})^2/2\sigma] \quad (2)$$

where $D_{RT}(R_i) = \underset{j=1,2,3,4}{min}\ d(c_i, c_j)$ is the minimal distance from i_{th} human enclosure center c_i to the four third-points c_j. R_i is i_{th} candidate enclosure, $diag$ is the length of diagonal line of image I, that is utilized to normalize the distance. ω_{vi} is the visual weight of R_i, and is defined by the study on visual perception which will be presented in the next.

Visual Balance

Visual balance is another key aesthetic principle which will be achieved when the gravity center of all objects is located at the center of the image. That means, the gravity center

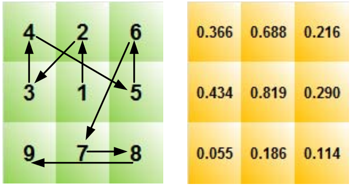


Fig. 3: Eye tracking results followed by [10]. Left: average scan path. Right: fixation map

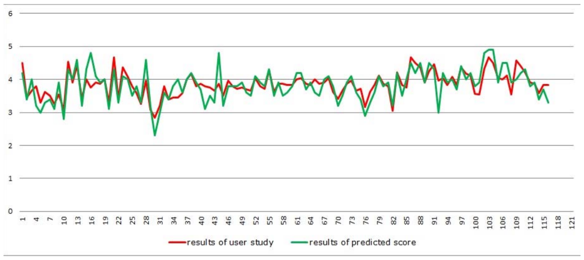


Fig. 4: Comparison of predicted score (green line) and average of subjective ratings (red line)

of i_{th} candidate enclosure should be located at its expected coordinate $\hat{c}_i$, so that

$$\left(\sum_{j=1}^{N_o}\omega_{sj}c_{oj}+\omega_{si}\hat{c}_i\right)/\left(\sum_{j=1}^{N_o}\omega_{sj}+\omega_{si}\right)=C_I \tag{3}$$

where N_o is the number of all salient objects in image I, c_{oj} is gravity center of j_{th} salient object, $\hat{c}_i$ is the expected gravity center of i_{th} candidate enclosure, ω_{sj} and ω_{si} are saliency values computed from j_{th} object and i_{th} candidate enclosure, respectively, and C_I is the center of image I. Thus, the expected enclosure center $\hat{c}_i$ can be calculated by:

$$\hat{c}_i=\left[C_I\times\left(\sum_{j=1}^{N_o}\omega_{sj}+\omega_{si}\right)-\sum_{j=1}^{N_o}\omega_{sj}c_{oj}\right]/\omega_{si} \tag{4}$$

Then the prediction score S_{VB} is defined as:

$$S_{VB}=\omega_{vi}100\times[exp(-\frac{D_{VB}}{diag})^2/2\sigma] \tag{5}$$

where $D_{VB}(R_i)$ is the distance between the i_{th} candidate enclosure center and i_{th} expected center $\hat{c}_i$.

To utilize attention distribution, we make use of the results from the eye tracking study reported in [10]. The fixation order and fixation map that depict attention distribution are shown in Fig 3. Depending on which part of the fixation map the candidate enclosure $c_i(x,y)$ belongs to, the visual weights can be defined. $Att(x,y)$ is an attractiveness value. Therefore, the visual weights for $R_i(c_i,s)$ can be calculated as $\omega_{vi}=Att(x,y)$, which aims to place people onto a more attractive region.

Aesthetic score prediction

Finally, we shall combine all parameters defined above to obtain the aesthetic score prediction function as:

$$S_{pre}=\frac{\exp(-\frac{{S_{oRT}}^2}{2{S_{oVB}}^2})S_{RT}+\exp(-\frac{{S_{oVB}}^2}{2{S_{oRT}}^2})S_{VB}}{\exp(-\frac{{S_{oRT}}^2}{2{S_{oVB}}^2})+\exp(-\frac{{S_{oVB}}^2}{2{S_{oRT}}^2})} \tag{6}$$

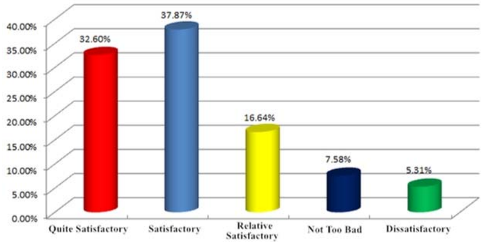


Fig. 5: Results of subjects evaluation. red: quite satisfactory; blue: satisfactory; yellow: relative satisfactory; dark blue: not too bad; green:unsatisfactory

where S_{oRT} and S_{oVB} are salient objects' rule of thirds score and visual balance score, respectively. They can be computed the same way by Eq.2 and Eq.5. The salient regions are obtained by graph-based method [11].

3. PHOTOGRAPHY SUGGESTION

Based on the aesthetic score prediction model, the ratings of each enclosure at different positions in various scales from the input image I can be predicted. The optimal one can be found by exhaustively searching through all possible enclosures with different positions and sizes. However, we have designed a hierarchical search scheme to obtain the optimal location of the enclosure efficiently.

3.1. View specific candidate enclosures generation

To start with, the enclosures that do not comply with specific view constraints will be discarded. We assume the input image I with size 640 × 426 or 426 × 640, which indeed is common for most mobile phones. We then slide a 196 × 128 window with step size as 32 over the given image. To obtain candidate enclosures, we adopt graph-based manifold ranking method [11] to obtain the salient region distribution so that these enclosures shall not overlaid on the salient object. In addition, according to the constraint to keep the enclosure inside the image, enclosures that across the image border will also be discarded. We also adopt the scheme in [12] to detect ground and sky regions Φ. This is intended to avoid selecting inappropriate regions that cannot be used to place a person.

3.2. Position and size optimization

3.2.1. Position finding

An enclosure's position $c(x,y)^T$ is dependent on the visual weights ω_v, size s and saliency value ω_s as defined above. Therefore, the positions can be solved first.

We carry out an iteration process to optimize the position. The initial point is the center $\vec{c}(x,y)^T$ of top enclosure searched by the last stage. We update the position by $\vec{c}_t(x,y)^T \leftarrow \vec{c}_{t-1}+\vec{\mu}$, where $\vec{\mu}$ is obtained by:

$$\vec{\mu}^*=\underset{\vec{\mu}'\in\{\vec{\mu}^{(n)}\}\cup\{0\}}{\arg\max}\; S_{pre}(\mathrm{R}(\mathrm{I};\tilde{c}_{t-1}+\vec{\mu}',s_{t-1})) \tag{7}$$

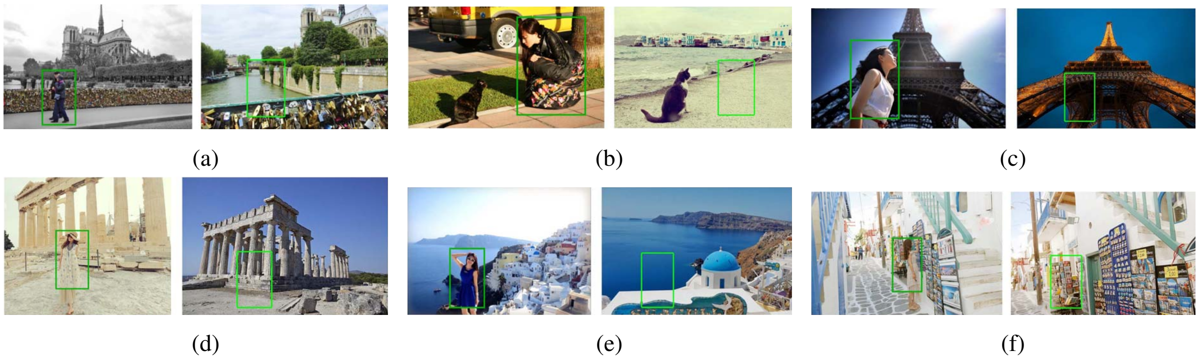


Fig. 6: Comparisons of hand labeled professional photos and suggested landscape photos with similar composition and background content

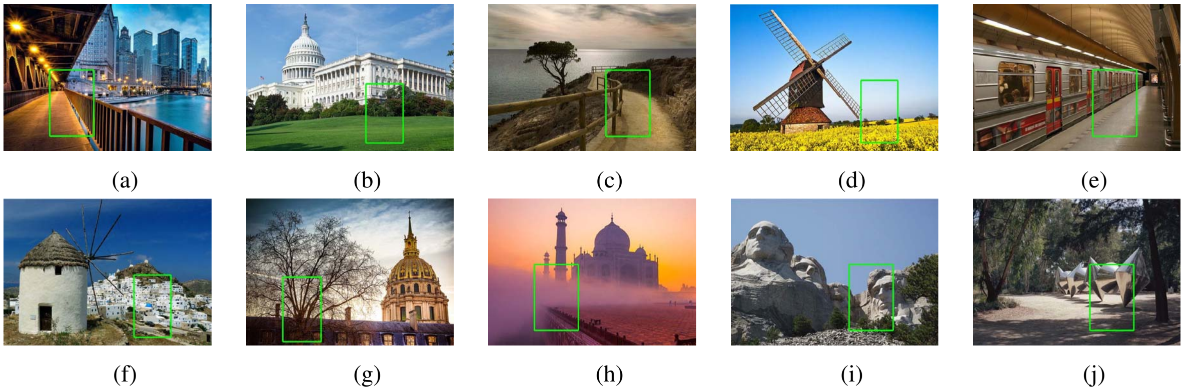


Fig. 7: Top 5 (fist line) and bottom 5 (second line) suggested photos rated by users

$\{\vec{\mu}^{(n)}\}_{n=1}^{N}$ is N $(N = 10)$ randomly generated vectors from a two-dimensional Gaussian distribution.

3.2.2. *Size update*

Once the position is confirmed, the size of enclosure can be obtained under the constraint that no blocking of any salient object shall occur. We iteratively re-scale the size of enclosure by $\pm 5\%$, until all constraints are satisfied.

4. EXPERIMENTS AND DISCUSSION

To validate the proposed system, we performed simulations on considerable amount of landscape photos from several hot spot landmarks and conducted subjective evaluations using 116 suggested images. 25 subjects including 9 females and 16 males with age ranging from 20 to over 30 are asked to rate the recommendation results. Among them, 2 subjects are professional photographers, 7 subjects have more than five years of photography experiences with single lens reflex cameras.

We compared the predicted score with subjective ratings to evaluate our aesthetic score prediction model. The comparison (Fig 4) shows that the predicted score by the proposed scheme is quite consistent with the subjective assessment. As shown in Fig 5, the satisfaction rate (score$\geq$3) reaches 87.11%, while the unsatisfactory rate (score$\leq$2) is at only 5.3%. Furthermore, we compared some professional photos with hand labelled bounding box with the suggested photos that have similar composition and background content, as shown in Fig 6. As can be seen from Fig 6, under similar circumstances, the recommendations from the proposed scheme are highly correlated to the composition of the professional photographers. We have also shown in Fig 7 top five and bottom five suggested photos rated by subjective evaluations. For these unsatisfactory suggestions as shown in Fig 7f, 7g, 7h, users consider that it is difficult to place human in these positions. In Fig 7i, 7j, the enclosures are blocking part of the salient objects due to inaccurate salient region detection. In summary, the proposed scheme is indeed able to find satisfactory enclosure for most scenery pictures to place a person in the scene and can be useful for novice photographers.

5. CONCLUSION

In this paper, we proposed a photography suggestion scheme to assist mobile users to compose high visual quality photos with human in the scene. An aesthetic score prediction model has been constructed to measure the visual quality. A study on professional photos is also conducted to define a proper size for enclosure. To provide a view specific solution, the proposed scheme leverages saliency and geometric detection results to represent composition features. Finally, an efficient hierarchical search has been developed to obtain an optimal enclosure effectively. Extensive experiments have been performed and the results have shown that the proposed photography suggestion system is promising in guiding novice photographers to take professional like appealing photos.

References

[1] W. Yin, T. Mei, C. W. Chen, "Crowdsourced learning to photograph via mobile devices," in Proceedings of IEEE International Conference on Multimedia & Expo, pp.812-817, 2012.

[2] W. Yin, T. Mei, C. W. Chen, and S. Li, "Socialized mobile photography: learning to photograph with social context via mobile devices," IEEE Transactions on Multimedia, vol.16, no.1, pp.184-200, 2014.

[3] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in Proceedings of European Conference on Computer Vision, pp.288-301, 2006.

[4] P. Obrador, L. S. Hackenberg, N. Oliver, "The role of image composition in image aesthetics," in Proceedings of IEEE International Conference on Image Processing, pp.3185-3188, 2010.

[5] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in Proceedings of ACM Multimedia, pp.271-280, 2010.

[6] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1657-1664, 2011.

[7] W. T. Chu, Y. K. Chen, and K. T. Chen, "Size does matter: how image size affects aesthetic perception?," in Proceedings of ACM Multimedia, 2013.

[8] B. Cheng, B. Ni, S. Yan, and Q. Tian, "Learning to photograph," in Proceedings of ACM Multimedia, pp.291-300, 2010.

[9] W. Yin, T. Mei, C. W. Chen, "Assessing photo quality with geo-context and crowdsourced photos," in Proceedings of IEEE International Conference on Visual Communications and Image Processing, pp.27-30, 2012.

[10] M. Smit. "Optimising advertising roi on print in a rapidly changing economic and media landscape - the contribution of eye tracking now and into the future," in Proceedings of Southern African Marketing Research Association Conference, 2012.

[11] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.3166-3173, 2013.

[12] D. Hoiem, A. A. Efros, and M. Hebert, "Geometric context from a single image," in Proceedings of IEEE International Conference on Computer Vision, pp.654-661, 2005.