AESTHETIC COMPOSITION REPRESENTATION FOR PORTRAIT PHOTOGRAPHING RECOMMENDATION

Yanhao Zhang¹, Xiaoshuai Sun¹, Hongxun Yao¹, Lei Qin², Qingming Huang²,³

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
²Key Lab of Intel. Inf. Proc., Inst. of Comput. Tech., Chinese Academy of Sciences, Beijing, 100190, China
³Graduate University of Chinese Academy of Sciences, Beijing, 100190, China
{yhzhang, xssun, h.yao}@hit.edu.cn {lqin, qmhuang}@jdl.ac.cn

ABSTRACT
In this paper, we present an intelligent portrait photographing framework for automatically recommending the suitable positions and poses in the scene of photography taken by amateurs. By analyzing aesthetic characteristics features, we propose a solution by constructing aesthetic composition representation which covers the attention composition and geometry composition to identify the underlying technique of professional photographer. First, we extract the attention composition feature of the professional photo by utilizing a visual saliency model. Then, a geometry composition feature is also presented to learn the spatial correlation. Finally, composition rules are applied to make appropriate pose and position. Experiments show our aesthetic composition representation performs well for portrait photographing recommendation.

Index Terms— Attention and Geometry Composition, Aesthetic Recommendation, Composition Rules.

1. INTRODUCTION
Aesthetic photographing and photo quality assessment are popular and extremely challenging topics. Previous works of the aesthetic based assessment approach dealing with the content-aware photo management problems mainly evaluate the quality of photos according to photography principles. Many interesting applications including perceptual photo quality assessment, photo view recommendation and quality based photo re-ranking have drawn great attention in multimedia research community. C. Li et al. [1] propose a framework of automatically evaluating the aesthetic quality on photos with faces. Low-level image composition features are used to build aesthetic classifier for aesthetic appeal assessment in [2]. Some works attempt to make personalized image tag recommendation [3] and photo selection towards aesthetic [4]. Based on aesthetic-related feature analysis, the above applications show great potential to improve the perceptual photo quality and guide the users to take photos that follow the aesthetic rules.

As the dramatic increase of the digital camera devices, many people try to take photos by themselves. However, most users take portrait photos following their casual feelings. For most amateurs, when taking photos in a great landscape and landmark, they always ask themselves how to pose and where to stand in the beautiful scene. From a professional point of view, the photo composition rules that consider the arrangement of the spatial context contribute a lot to the quality of the photos. Experienced photographers adhere the aesthetic elements and rules to the artworks, which makes photos more visually appealing.

Professional photos were taken using professional photography techniques and consistent with aesthetic composition rules. If the users get the recommendations from the similar professional work, it will fully satisfy users’ needs via their amateur photographing. Moreover, knowledge of the professional photos could help the users to shoot better pictures and enrich the users’ experience during photographing. Inspired by this intuition, we propose a novel intelligent photography approach to recommend the suitable pose and position of a portrait in a given camera view scene. Different from the previous works considering the aesthetic features of the entire image, ours focus on aesthetic composition representation by emphasizing the attention and geometry composition of the photo which is closely related to the aesthetic photographing rules. Similar with [5], the framework of the recommendation approach we proposed (shown as figure 1) can equip the digital cameras with smart function which could help the user for shooting near professional photos.
2. METHOD OVERVIEW

Figure 2 gives a pipeline of the proposed framework. To accomplish the aesthetic purpose of photographing, the crawled professional online sharing photos are first collected to form a reference photo dataset. Then, two methods are presented to obtain the aesthetic spatial composition representation from the professional photos based on attention and the geometry composition feature respectively. When a query photo without portrait is presented, we retrieve the reference photos in the database using the above mentioned composition representation. Finally, we select one single reference photo from the candidates adaptively with suitable pose and position. As the recommendation is produced using professional photographs, it is consistent with the classical composition rules and guide the users to take satisfactory photos.

3. AESTHETIC COMPOSITION REPRESENTATION

In this section, we discuss two composition features to build aesthetic composition representation for reference photo recommendation. The goal is to select the professional photo which shares the largest aesthetic composition similarities with the query image. At beginning, we use human pose estimation method for still images in [9] to make pose estimation automatically on each professional photo. We re-annotate poses and positions of the photos with bounding box. Figure 3 shows some examples from the dataset with a ground truth aesthetic position and pose. The annotations associated with each professional photo are presented for recommendation.

3.1. Attention composition feature

3.1.1. Estimating attention distribution

Visual saliency based techniques are adopted to obtain a suitable estimation of spatial distribution of the subject region in the photographs. As we know, saliency maps with or without main subject will be very different. However, the main subject in the saliency is also a component of the spatial attention composition. In order to represent scene structure without stressing the portrait, we design a decaying exponential function to weaken the magnitude of saliency but preserving spatial attention distribution by Eq.1. We use PQFT method [6] to obtain initial saliency map $S$ to express the salient region distribution.

$$f(S(x, y)) = \begin{cases} 1 & S(x, y) < \bar{S} \\ \exp(-\beta S(x, y)) & S(x, y) \geq \bar{S} \end{cases}$$  

where $S^d$ refers to the decay saliency map, $\bar{S}$ is the average value of $S$. When $S(x, y)$ is greater than $\bar{S}$, we weaken the high magnitude to make the distribution uniform. By this way, decay saliency map can be used to represent the attention composition of the professional photos.

3.1.2. Attention composition feature extraction

To make it possible to match in the feature space, we quantize the decay salient map of each photo $S^d_i$ into a 1024-dim vector. In order to maintain all the spatial information, we resize the photo to $32 \times 32$ scale $S'(x, y)$. For each image $I_i$ in professional photo set $\Gamma$, a 1024 vector $U_i$ is formed by concatenating the value of $S'(x, y)$.

3.1.3. Hierarchical K-means clustering and matching

We need to find the nearest-neighbors for the query images in the 1024-dim low level visual feature space. Because of the high dimension, we choose to use the Hierarchical K-means method to speed up the search process. We first divide the space into K clusters by K-means method ($K=8$ in the experiment). Then, the query image is assigned to the closest cluster. We get the nearest results in the assigned cluster according to $D_{att}(I) = \text{dist}(U_q, U_i)$ which is the attention distance between query image and photo in the dataset. Figure 4 visualizes the matching process.

![Fig.4. Photo matching based on hierarchical K-means.](image-url)
3.2. Geometry composition feature

Besides attention composition, geometry composition is also an important cue to measure the scene structure composition. It provides additional discriminative amendment for aesthetic composition representation.

3.2.1 Geometry composition feature extraction

For the professional photos set \( \Gamma \), we attempt to use hill-climbing algorithm for color image segmentation [7]. All of the photos are assumed to have less than 10 regions of segmentation. So we set the total region number 10 for normalization. For each region of the input photo, we use \( C= (\text{norm}(x), \text{norm}(y), \text{norm}(\text{area}(x,y))) \) as its 3-dim compositional feature, where \( \text{area}(x, y) \) is the region area of segmentation, \( \text{coordinate}(x, y) \) is the centroid of the area. \( \text{norm}() \) is a function that normalizes the range of the value. Therefore, the 30-dim feature of the geometry composition is as follows:

\[
V_i = (\text{norm}(x_{i1}), \text{norm}(y_{i1}), \text{norm}(\text{area}(x_{i1}, y_{i1})), \ldots, \\ \text{norm}(x_{i10}), \text{norm}(y_{i10}), \text{norm}(\text{area}(x_{i10}, y_{i10})))
\]  

(2)

We set the unexacting region by default parameters \((x, y, \text{area})=(0.5, 0.5, 0)\). We take geometry distance between the query image and each photo \( D_{geo}(I_i) = \text{dist}(V_q, V_i) \) to rank the nearest-neighbor. The matching results between the query and database images with geometry composition similarity are illustrated in Figure 5.

![Matching results of geometry composition features.](image)

3.3. Rules based reference photo selection

Adaptive reference photo selections based on aesthetic composition representation are presented by Eq.3 and Eq.4:

\[
I_{can} = \arg \min_{i \in \Gamma} (\alpha I_{geo}(I_i) + \beta I_{config}(I_i))
\]  

(3)

where \( I_{can} \) denotes the candidates, we use \( \alpha = 0.6, \beta = 0.4 \) in the experiment, since different weight depends on personal preference. By Eq.3, 8 candidate photos are selected. We associate the searched aesthetic candidates to the well-defined composition rules of professional photos. The rules of thirds and the golden ratio segmentation are the most well-known photograph composition principles. The ideas are to place the main subjects on the certain place of the photograph. Practically, we adapt four stress points \((S', \ldots, S')\) from rules of thirds and four stress points \((S', \ldots, S')\) from golden ratio as well [2]. So we get eight points to determine the visual subject center aesthetic appeal level. We measure the aesthetic score by Euclidean distance between pose center \((x^0)\) and the eight stress points. Reference photo is to recommend with the minimum aesthetic score as Eq.4:

\[
\text{Score} = \min(\text{dist}(x^0, s_i)), i = 1 \text{ to } 8;
\]

\[
I_{can} = \arg \min_{i \in \Gamma} (\text{Score}(I_i));
\]  

(4)

where \( I_{can} \) is the selected reference photo. We evaluate the usefulness of aesthetic composition representation in next section.

4. EXPERIMENTS AND DISCUSSION

4.1. Data Collection

Most of the previous works on this topic collect private dataset to evaluate their performance. In order to guarantee the underlying professional spatial arrangements of the mined photos, we crawled photos from a professional photo contest website, namely photo.net [8]. Our dataset consists of 232 photographs, mainly from editorial and fashion photographers’ works. The test data are 50 images taken by cameras, mainly captured in the scenes like campus, park, landscape and landmarks. Images are all resized to \( 480 \times 640 \). We apply the 50 test images to the recommend approach for our experiment.

4.2. User Study

We conducted an effective evaluation of the proposed method as follows. After getting the recommendation photos, 10 participants (3 female, 7 male) from 23 to 30 years old help to give the evaluation of the pose and position. Each measures the recommendation results of 50 images for five satisfaction levels. For the procedure of evaluation, each person is asked to mark the place and pose they want to capture in the images firstly. Then, each user gives the satisfaction level by comparing with recommendation to make it subjective. The statistics of the user studies on satisfaction degree are illustrated in Figure 6. As we can see, the satisfaction rate (above Satisfactory) is 57.6%. Also, we show some cases of the success and the failure photos that the approach recommends in term of the user feedback in Figure 7.

![Statistics of the satisfactory level. 10 participants vote for 50 recommend results. We summarize the results of five level (“Quite Satisfactory”, “Satisfactory”, “Relative Satisfactory”, “Not too bad”, “Dissatisfactory”) were (11.6%, 46%, 28%, 12.4%, 2%).](image)
4.3. Experimental results analysis
The first 3 examples (a), (b), (c) illustrated in Figure 7 show that, in these cases, when the photo’s attention composition is quite similar with the geometry composition in simple background, the recommendation results are quite satisfactory. For the cases (d), (e), (f), even through the backgrounds of the queries and the returned recommendation are quite different, hardly related in any way, the results is satisfying. This is because in this circumstance, the reference photo mainly depends on the attention composition similarity in the background rather than the geometry composition similarity, and the users are favored by the attention composition of the scene which coincides with that of the reference photo. For some failure cases (g), (h), (i), our matching method based on attention and geometry composition similarities performs well in term of capturing the composition features between the query and reference in these cases. However, the results of the pose and position are difficult for the users to achieve, because we don’t take the spatial and semantic contextual information into consideration, for instance people can’t stand in water or fly in the sky. These examples demonstrate if we add semantic context information into our model, we may obtain better results. While it is beyond the scope of this paper, we may explore it in the future.

5. CONCLUSION
In this paper, we construct aesthetic composition representation by utilizing attention and geometry composition features. By means of that, a novel pose and position recommendation method is proposed for portrait photographing. The representation captures the scene structure of the photos and efficiently covers overall composition characteristics. We conduct adaptive reference photo selection with composition rules. User feedback is carried out on a collected ground truth data. Experiments show that our representation lead to promising results.

6. ACKNOWLEDGMENTS
The work was supported by the National Natural Science Foundation of China (No. 61071180) and Key Program (No. 61133003).

7. REFERENCES

Fig.7. Experimental results of the recommendation pose and position by the proposed method. In each case, first row: input images; second row: reference photo. The annotated color pose are position are recommended from reference photo to input image. Bottom right: the average statistics feedback of the cases group.