POSE-BASED COMPOSITION IMPROVEMENT FOR PORTRAIT PHOTOGRAPHS

Xiaoyan Zhang[1], Kap Luk Chan[2], Martin Constable[3]*

[1]College of Computer Science & Software Engineering, Shenzhen University, China 518060
[2]Tolendata Singapore R&D Centre Private Limited, Singapore, [3]RMIT University, Vietnam
Email: xyzhang15@szu.edu.cn, kapluk_chan@hotmail.com, martin.constable@rmit.edu.vn

ABSTRACT

This paper studies the composition of portrait paintings and develops an algorithm to improve the composition of portrait photographs. The study of portrait paintings shows that placement of the face and the figure in portrait paintings is pose-related. Based on this observation, this paper develops an algorithm to improve the composition of a portrait photograph by learning the placement of the face and the figure from an example portrait painting. The example portrait painting is selected based on the similarity of its figure pose to that of the input photograph. This similarity measure is modeled as a graph matching problem. Finally, space cropping is performed using an optimization function. Experimental results and a user study demonstrate that the proposed pose-based improvement is preferred more than rule-based methods.

***Index Terms*—** Composition; Pose; Portrait

1. INTRODUCTION

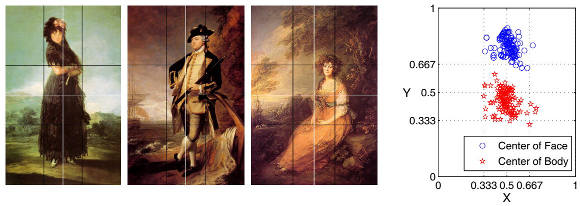
Composition being one of the important aesthetic aspects, requires attention to be improved in photographs. In this paper, we address a specific genre of photographs: portraits, which are photos focusing on the depiction of a person. In portrait paintings, the size and the positions of the figure and the face are specially organized by artists to draw all the attention onto the subject. In contrast, in snap-shot style photographs by amateurs, those are often not well organized. When selecting personal photos for sharing in social networks or for printing for home display and photo album, we often discard photos in which the face is too small or where the subject occupies or intersects the corners or edges of the image. In this situation, the composition can be improved first before sharing or printing. Motivated by this, we propose to improve the composition of portrait photographs by optimizing the size and position of the figure in the image.

Currently, rule-based methods have been developed to improve the composition of portrait photographs [1, 2, 3]. They optimize the composition based on rule constraints (e.g. rule of thirds, one of fifth or center, etc.). The main feature they considered is the location of the face. These rules can produce photographs more likely to be pleasing than placing the face in an arbitrary location. However, no absolute rule can ensure good composition in images. In portrait paintings, artists convey the personality of the subject through attention to their facial expression and the bodily pose [4]. When planning a portrait, the artist first carefully finds a natural pose for the subject [5]. Then, the artist studies the subject for a facial expression that conveys his understanding of the subject's significance. Finally, the artist will compose the portrait by selecting the best view [5]. Hence, we can say that the composition in a portrait is pose-related, not simply rule-based. Inspired by this, this paper proposes to improve the composition of a portrait photograph according to the pose of the figure. An example portrait painting in which the figure has similar pose with that in the photograph is selected to guide the improvement of the photograph. Based on our knowledge, this is the first work that improves composition of portrait based on pose and uses example paintings to guide the composition improvement.

Thanks to Institute for Media Innovation, Nanyang Technological University, Singapore, for providing the Ph.D. grant to complete part of the work. The authors wish to acknowledge the financial support from Chinese Natural Science Foundation (61602313), Shenzhen University research funding (2016051), and Tencent "Rhinoceros Birds"-Scientific Research Foundation.

2. RELATED WORK

Composition improvement for portrait. Composition improvement is a popular topic for researchers. However, the composition improvement for portrait is seldom discussed. The existing methods are generally all rule-based. Zhang et al. [1] presented an automatic photo-cropping algorithm based on constraints on the face location, the center of the region of interest (ROI), and the areas of the face and ROI. The rule of thirds, diagonal dominance, visual balance, and salient-region size were used in [2] to formulate an optimization function for composition improvement. Li et al. used the method in [2] to enhance the aesthetics of photographs with a face [3]. Although learning-based methods [6, 7] were effective to learn composition rules, they did not focus on portrait photographs. Based on our study on paintings, the composition of portrait is not simply rule-based, and it is pose-related. Different from the methods above, our method improves the composition of a portrait photograph based on the pose.



(a) (b) (c) (d)

Fig. 1. (a)-(c) Paintings; The dark lines are the one-of-third lines and the white lines are the vertical and horizontal central lines; (d) Locations of figures in 120 full-body paintings.

Human pose retrieval. Ferrari et al. originally proposed a human pose retrieval method based on pose estimation [8]. The full version was further published in [9]. Human pose estimators were used for the pose estimation. It was capable of estimating upper and full body poses. In order to improve the estimation performance, body part detectors were used for the pose estimation and further applied for pose retrieval [10]. Although it was effective, only upper body pose was estimated. In our work, the full body pose needs to be estimated too. Therefore, we extract the pose based on the work in [8].

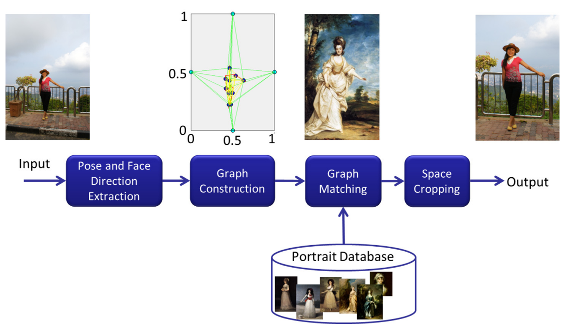


Fig. 2. The framework of the proposed method.

3. COMPOSITION IN PORTRAIT PAINTINGS

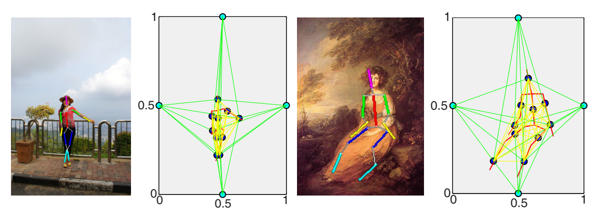
Portrait paintings collected for this study are chosen from artists in the 17-18 centuries. They painted a lot of portraits with outdoor background and they usually maintain a high dynamic range which is broadly similar to that of photographs. In total, 300 portrait paintings are collected to form the database. 180 of them are half-body portraits and 120 are full-body portraits. The half-body portrait paintings are classified to small face (118 paintings, face/image height ratio < 0.2) and big face (62 paintings) half-body portraits.

In portrait paintings, the pose of the figure and the placement of the face are commonly considered together [11, 12]. Three example paintings with figures in different poses are shown in Fig. 1. The placements of the faces and figure bodies in the three paintings are clearly different as the figures are in different poses. The centers of faces and centers of the bodies of 120 full-body paintings are drawn in Fig. 1(d). It shows that the location of the face is not limited to the rule of thirds and the placements of the face and body have large variances. The various placements of faces and figure bodies in paintings indicate that the aesthetics of the portrait is not simply rule-based and is pose-related.

4. COMPOSITION IMPROVEMENT METHOD

The framework of the proposed method is in Fig. 2. An example painting with figure in the pose similar to that of the figure in the photograph is selected as the reference to guide the composition improvement of the photograph. Here, pose represents the body pose together with the face tilt direction and is extracted using the method in [8]. The face rotation angle is extracted separately as the face direction by the method proposed in [13]. The location of the figure is also considered in the example painting selection to make sure that the photograph can be cropped to have a similar location arrangement as the selected example painting.

Because the pose and location of the figure can be described as the relative relationships of body parts and body parts with the boundaries of the image, therefore, a graph $G = (V, E)$ is used to model the relative relationships, where V is the node of the graph and E is the edge that links the connected nodes. Two example graphs are shown in Fig. 3. There are 10 body nodes for full-body portrait and 6 for half-body portrait.



(a) (b) (c) (d)

Fig. 3. Graphs (b) and (d) for a photograph (a) and a painting (c); The nodes in dark blue are the body part nodes. The nodes in cyan are the boundary nodes. The red lines are the body skeleton. The yellow lines link the body nodes and the green lines link the head, lower arm and lower leg nodes with the boundary nodes.

After constructing the graph, the example painting is selected by conducting graph matching. Then, space cropping is performed by using an optimizing function.

4.1. Graph matching

After constructing the graph to describe the pose and location of the figure, the example painting selection can be achieved by a graph matching based on a similarity measure. Given the input graph G_I and the graph of the k-th painting G_k in a database, the graph matching is formulated as

$$\max_k S(G_I, G_k) \tag{1}$$

While the correspondences of nodes and edges of two graphs are fixed, the similarity of two graphs can be defined as:

$$S(G_I, G_k) = \sum_i S_i^v(G_I, G_k) + \sum_j S_j^e(G_I, G_k) \tag{2}$$

where, $i = 1, ..., N$, N is the number of nodes, and $j = 1, ..., M$, M is the number of edges. S^v is the similarity of nodes and S^e is the similarity of edges. The definition of similarities of edges and nodes is based on their descriptors.

The descriptors of a body node v^b are the location and orientation of the corresponding body part. The descriptors of the edge e^b linking the body nodes is the relative location and orientation of the two corresponding body parts. In the pose estimation by [8], the posterior marginal distribution of the position (including location and orientation) of each body part is estimated. Based on it, the location and orientation descriptors of a body node and the relative location and orientation between body nodes are calculated. Details of the calculation of these descriptors can be found in [8].

Specifically, the face node is described by the face direction. It is described by 13 quantized degrees from 90° to -90° [13]. The descriptor of the boundary node v^d is the boundary center. The descriptor of the edge e^d that links the body node with the boundary is its distance in the x or y dimension.

The combined Bhattacharyya similarity is used to calculate descriptor similarities of body nodes v^b and edges e^b [8]. The similarity of two face directions θ_1 and θ_2 is measured as

$$S_F(\theta_1, \theta_2) = e^{-(\frac{\theta_1 - \theta_2}{\sigma})^2} \tag{3}$$

σ controls the sensitivity to the difference of face directions.

The constraint on the distance of the body node with the boundary node is to make sure that there is enough space in the photograph for cropping to achieve the composition of the example painting. While the distance in the photograph is bigger than that in the painting, the similarity should be the maximum value, otherwise the similarity is decreasing in proportion with the increase of their difference. Based on this definition, given a standard variance σ_d, the similarity of an e^d edge from the input graph with the corresponding e^d edge from a painting graph with descriptors d_I and d_k is

$$S^{e^d} = \begin{cases} 1 & \text{if } d_I - d_k \geq 0 \\ e^{-(\frac{d_I - d_k}{\sigma_d})^2} & \text{if } d_I - d_k < 0 \end{cases} \tag{4}$$

4.2. Space cropping

Space cropping is used to assign a similar location for each body part of the photograph with that of the example painting by finding a cropping window. Because sizes and aspect ratios of the figures are different in the photograph and the example painting, we cannot directly crop the photograph by giving the same space around the figure as in the example painting. Hence, the space cropping is formulated as an optimization problem. The energy function is defined as

$$E = \alpha \frac{1}{N^b} \sum_i^{N^b} \|v_i - v_i^r\| + (1 - \alpha)\|v_{face} - v_{face}^r\| \tag{5}$$

The first term is the pose constraint. v_i is the target location for the body part i and v_i^r is the reference location from the example painting. N^b is the number of body parts.

Due to the importance of the face in a portrait, a face constraint is added to the energy function as the second term to prevent the face from shifting too much from the location in the painting. v_{face} is the target location of the face and v_{face}^r is the reference location. The pose and face constraints are linearly combined by a constant scale α.

The cropped rectangle is represented by the left top corner coordinates (l, t), width (w) and height (h). Given a specified aspect ratio a, the height can be calculated as $h = a \times w$. Therefore, the optimization of the energy function can be reformulated as finding a vector of (l, t, w). The particle swarm optimization (PSO) [14] method is used to seek the optimal solution by globally searching the minimum candidate of (5).

5. EXPERIMENTS

The algorithm was implemented using MATLAB on a PC with an Intel 2.67GHz processor and 8GB RAM. For a photograph in dimension of 1000×750, it took around 0.95 seconds to obtain the optimization result using PSO. The parameters used in our implementation were determined empirically as $\sigma = 45$, $\sigma_d = 0.3$, and $\alpha = 0.4$. Based on the study of aspect ratios of portrait paintings, the aspect ratios for full-body and small face half-body portraits were defined as $a = 1.5$, and for big face half-body portraits, it was $a = \frac{4}{3}$. The test portrait photographs were collected from the MIT-Adobe FiveK Dataset [15] and our personal collections.

Fig. 4 shows the cropping results of photographs with figures in different poses. The poses of the figure in the selected example paintings are similar to those of the corresponding input photographs. In composition improvement results in Fig. 4, the locations of the figures are similar to those in the corresponding example paintings. These figure locations serve the requirement of their poses.

5.1. Comparison with rule-based methods

To evaluate the effectiveness of our pose-based composition improvement method, it is compared to two rule-based methods designed for photographs with a face. One method is the ACDP method proposed in [1]. The other method is the ABPE method developed in [3]. Composition improvement

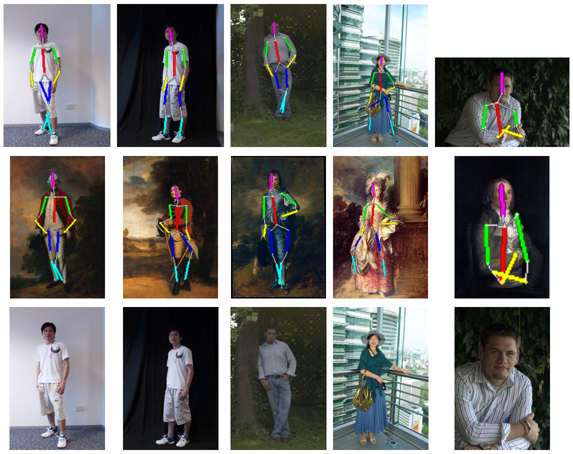


Fig. 4. Photographs (top) and selected example paintings (middle) with extracted poses; Bottom: cropping results.

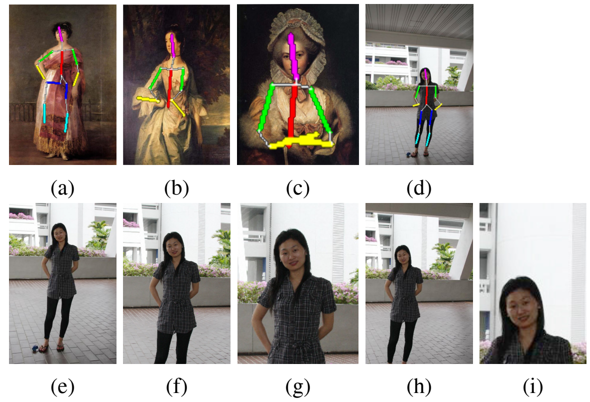


Fig. 5. (a)-(c) Example paintings and (d) input photograph with extracted poses; (e)-(g) Results by our method based on (a)-(c); (h) ACDP method result; (i) ABPE method result.

results for a photograph are shown in Fig. 5. The result in Fig. 5(h) is produced by the ACDP method using a small face template. The template is selected based on the face size of the original photograph. In the ACDP method, the ROI is used to constrain the cropping. The ROI is detected by saliency detection. The detected ROI may lose some important body parts. In Fig. 5(h), the feet are cropped out which leaves the figure unbalanced in the image. The ABPE method [3] only considers the face region as important information, hence in the result Fig. 5(i) only the face and shoulder are visible. Additionally, the face is placed at the bottom one third point while leaving half of the space on the top for background. It is not reasonable considering the visual appearance of the whole image. For a full-body portrait photograph, our method can crop it to a full-body portrait (Fig. 5(e)), a half-body portrait with small face (Fig. 5(f)), or a half-body portrait with big face (Fig. 5(g)). The results by our method are more encouraging than those by rule-based methods.

5.2. User study

In order to significantly evaluate the advantage of our method, a user study was conducted to compare composition improvements by our method with those by ACDP [1] and ABPE [3] methods. 20 test photographs were randomly selected for this user study. The participants connected to the user study webpage designed by Google Form using their own computers.

The results obtained by our method and one of other methods were shown as a pair side by side in randomly generated order. The participants were asked to answer a question “Which is better composed and more suitable for printing for home display or making a photo album?”. They could choose “Left”, “Right”, or “They look the same” as a response. There were a total of 18 participants aged from 20 to 35. 11 were male and 7 were female. For the comparison with each method, 360 responses were collected. The user study result is summarized in Table 1. It shows that our results are preferred by a significantly higher percentage of responses than those by other methods. It is not surprising that our results are preferred more in comparison with those achieved using ACDP method, because ACDP method is easily disturbed by background objects. The positive responses to our results are 4 times greater than those to ABPE method. The fewer responses to ABPE method is due to that it forces the face to a one third point without considering other body parts. The face sometimes is placed on the bottom one third point (see Fig. 5(i)) and results in leaving large space for the background. This demonstrates the importance to consider the face and the body together in composition improvement.

Table 1. User study result for the comparison.

Our vs ACDP	Our	ACDP	Same
	68.3%	22.2%	9.5%
Our vs ABPE	Our	ABPE	Same
	81.4%	16.4%	2.2%

6. CONCLUSION AND DISCUSSION

This paper proposes to improve the composition of a portrait photograph based on an example portrait painting. The example painting is selected based on figure and face poses, and the location of the figure, which are modeled as a graph. Space cropping is formulated as an optimization problem to improve the composition. Experimental results and a user study show that our method performs better than rule-based methods. However, our method relies on the accuracy of the pose estimation. In the future work, we will develop an interactive method to correct wrongly estimated body parts. Meanwhile, in the example painting selection, the context information of the background will be considered and we will extend the proposed method to use professional photographs as examples to guide the composition improvement.

7. REFERENCES

[1] Mingju Zhang, Lei Zhang, Yanfeng Sun, Lin Feng, and Weiying Ma, "Auto cropping for digital photographs," in *Proc. of the IEEE Conference on Multimedia and Expo*, 2005, pp. 438–441.

[2] Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, pp. 469–478, 2010.

[3] Congcong Li, Alexander C. Loui, and Tsuhan Chen, "Towards aesthetics: a photo quality assessment and photo selection system," in *Proc. of the ACM international conference on Multimedia*, 2010, pp. 827–830.

[4] Willem H. Vanderburg, *Our Battle for the Human Spirit: Scientific Knowing, Technical Doing, and Daily Living*, University of Toronto Press, 2016.

[5] Wendon Blake, *Oil Portraits: Step by Step*, Courier Dover Publications, 1998.

[6] Luming Zhang, Mingli Song, Qi Zhao, Xiao Liu, Jiajun Bu, and Chun Chen, "Probabilistic graphlet transfer for photo cropping," *IEEE Transactions on Image Processing*, vol. 22, pp. 802–815, 2012.

[7] Bingbing Ni, Mengdi Xu, Bin Cheng, Meng Wang, Shuicheng Yan, and Qi Tian, "Learning to photograph: A compositional perspective," *IEEE Transactions on Multimedia*, vol. 15, pp. 1138–1151, 2013.

[8] Vittorio Ferrari, Manuel Marin-Jimenez, and Andrew Zisserman, "Pose search: Retrieving people using their pose," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1–8.

[9] Marcin Eichner, Manuel Marin-Jimenez, Andrew Zisserman, and Vittorio Ferrari, "2d articulated human pose estimation and retrieval in (almost) unconstrained still images," *International Journal of Computer Vision*, vol. 99, pp. 190–214, 2012.

[10] Nataraj Jammalamadaka, Andrew Zisserman, and C. V. Jawahar, "Human pose search using deep poselets," in *Proc. of International Conference on Automatic Face and Gesture Recognition*, 2015.

[11] Harry Berger, *Fictions of the Pose: Rembrandt Against the Italian Renaissance*, Stanford University Press, 2000.

[12] Dianne W. Pitman, *Bazille: Purity, Pose, and Painting in the 1860s*, Penn State Press, 1998.

[13] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation and landmark estimation in the wild," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2879–2886.

[14] James Kennedy and Russell Eberhart, "Particle swarm optimization," in *Proc. of the IEEE International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948.

[15] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 97–104.