Pose-based Composition Improvement for Portrait Photographs

Xiaoyan Zhang, *Member, IEEE,* Zhuopeng Li, Martin Constable, Kap Luk Chan, *Member, IEEE,* Zhenhua Tang, and Gaoyang Tang

***Abstract*—This paper studies the composition in portrait paintings and develops an algorithm to improve the composition of portrait photographs based on example portrait paintings. A study of portrait paintings shows that the placement of the face and the figure 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. This 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 to assign a similar location for each body part of the figure in the photograph with that of the figure in the example portrait painting. The experimental results demonstrate the effectiveness of the proposed method. A user study shows that the proposed pose-based composition improvement is preferred more than rule-based methods and learning-based methods.**

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

I. Introduction

COMPOSITION, being one of the important aesthetic aspects that influence visual quality, requires attention to be improved in photographs. Photographs with a different subject focus need to be composed in different ways. In this paper, we address a specific genre of photographs: portraits, which are photos focusing on the depiction of a person. In portrait paintings, where the placements of the figure and the face are the most important elements in the composition [1], 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 taken by amateurs, the size of the figure and the spatial composition 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. This we might do despite the face being presented in a pleasing manner. 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. Rule-based methods 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 photographs where the face has been placed in an arbitrary position. However, no absolute rule can ensure good composition in all images. In portrait paintings, artists convey the personality of the subject through attention to facial expression and the body pose [2]. When planning a portrait, the artist first carefully finds a natural pose for the subject that looks active [3]. Then, the artist studies the subject for a facial expression that satisfactorily conveys his understanding of the subject's significance. Finally, the artist will compose the portrait by selecting the best view [3]. 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 from a database of portrait paintings to guide the improvement of the photograph. A graph model is developed to describe the pose. The example painting is selected by searching a matching graph. According to the location of each body part in the selected example painting, space cropping is then conducted for the photograph by optimizing an energy function. Based on our knowledge, we are the first that proposes to improve the composition of portrait based on the pose of the figure and guide the composition improvement based on paintings.

The paper is organized as follows. Firstly, some related papers are reviewed in Section II. A study of portrait paintings is introduced in Section III. The observation from this study is that the placement of the figure in portrait paintings is pose-related. This provides the justification for the pose-based composition improvement of portrait photographs. Section IV describes the proposed pose-based composition improvement method. Experimental results and a user study are presented in Section V to illustrate the effectiveness of the proposed method. The conclusion is ordered in Section VI.

II. Related Work

Composition in aesthetics measure. Image composition takes an important role in image aesthetics. Most methods that measure the aesthetics of images use composition as part of their features. Rule of thirds is one popular rule used in the aesthetics measure [4], [5]. It is the rule that the center of region of interest should be placed on one of the power points (the four cross points of the horizontal and vertical one-of-third lines). The golden section or golden ratio also is used as an important composition rule in the aesthetics measure [6], [7]. In the aesthetics measure of portraiture, besides the rule of thirds [8], [9] and the golden section [10], the pose of the face [10], the ratio of the face area to the image size [11], [8], and the ratio of the body length to the face length [12] are also used as composition features. From the aesthetics measuring methods reviewed above, generally we can conclude that the size of the face area and the location of the face are important composition features for the aesthetics measure of portrait images. However, these methods only measure the location of face based on rules without consideration of the body pose.

Composition improvement. There are two objectives in composition improvement. One is to improve the composition for aesthetics. Another is to change the aspect ratio or downsizing the image while preserving the important image features for adaptive display. This technique is called retargeting. Although content-aware retargeting methods [13], [14], [15] are effective in composition correction, cropping is still one of the most favored methods comparing to others as it is less likely to introduce artifacts [16]. This paper use cropping to improve the composition of portraits for aesthetics.

Composition improvement is a popular topic for researchers. However, composition improvement for portraits is seldom discussed. Most of existing methods are rule-based methods. Zhang et al. [17] presented an automatic photo-cropping algorithm based on rule constraints on the face location, the placement of the region of interest (ROI), and the areas of the face and ROI. Based on the face count, ROI area count, and size of the face, photographs were categorized to 14 classes. A template was defined for each class based on the corresponding rules, e.g. rule of thirds, empty space, no middle, etc.. The rule of thirds, diagonal dominance, visual balance, and salient-region size were used in [18] to formulate the optimization function for composition improvement. Li et al. used the composition method in [18] to enhance the aesthetics of photographs with a face [19]. Although learning-based methods [20], [21], [22], [23], [24] were effective to learn composition rules, they did not focus on portrait photographs. Image cropping by training a crop window classifier from a training dataset consisting of photos before and after cropping by expert photographers was also proposed [25], [26]. It is a challenge to collect a big dataset consisting of photos before and after cropping for training. Therefore, the application of this method is limited. A human position recommendation system was developed for souvenir photography [27]. It was to decide the human location in the scene mainly based on the relative relationship between the human and the saliency feature of the background scenery. In contrast, our work does not change the location of the human in the scene and only conducts space cropping for the purpose of composition improvement.

Based on our study on paintings, the composition of portraits is pose-related. Different from existing methods, our method improves the composition of a portrait photograph based on the pose of the figure. Our previous work had no pose estimation correction and the performance was affected by pose estimation accuracy [28]. In this paper, both methodology and experiments have been improved with more details and analysis.

Human pose estimation and retrieval. Human pose estimation is challenging. It is not only suffered from external occlusion, self-occlusion, lighting inconsistency, but also human body is an articulate structure. Traditional approaches have relied on the aggregation of hand-crafted low-level features, which are then input to a standard classifier or higher level generative model. These engineered features are sensitive to variant in numerous deformations in the input space (such as variations in lighting). Recently, with the development of deep learning, human pose estimation has significantly progressed since the work of DeepPose by Toshev et al.[29]. Discriminative deep-learning approaches learn an empirical set of low and high-level features which are typically more tolerant to variations in the training set. However, incorporating priors about the structure of the human body into such networks is difficult. Tompson et al. [30] attempted to combine a Convolutional Network (ConvNet) part-detector with a part-based spatial-model into a unified learning framework to improve pose estimation performance. Marras et al. [31] emphasized the pairwise joint constraints in a three stage coarse-to-fine ConvNet architecture to improve the performance of the module. Ning et al. [32] proposed to inject external knowledge into the deep neural networks to guide its training process using learned projections that impose proper prior. Instead of adding constraints between joints, some researchers focus on the structure of the network itself and they aim to find a best end to end framework of human pose estimation. Wei et al. [33] proposed the convolutional pose machines, which consisted of a sequence of convolutional networks that repeatedly produced the location of parts or joints. Stacked Hourglass networks were used in [34], [35], [36] to improve the performance of the network. Multi-person pose estimation methods using deep learning were also developed [37], [38]. Although human pose estimation form static images have a significant progress using deep learning, there are still failure detections for partly occluded body parts (such as legs of a person wearing a skirt) and also for color disturbed parts. Additionally, the detection accuracy rate of body parts is obviously decreased in stylized images, such as paintings.

Ferrari et al. originally proposed a human pose retrieval method based on pose estimation to retrieve people using their pose [39]. The full version was further published in [40]. Human pose estimators were used for the pose estimation. It was capable of estimating upper body pose and full body pose. The descriptors for pose retrieval were probability distributions over possible part positions and orientations, and soft-segmentations of the parts. In order to improve the pose estimation performance, body part detectors trained by ConvNet were used for the body pose estimation and the pose feature from the ConvNet was applied for pose retrieval [41]. Although it was effective, only the upper body pose

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

Fig. 1. (a)-(c) Three paintings with different poses. The dark lines in the painting are the one-of-third lines and the white lines are the vertical and horizontal central lines. (d) The centers of faces and centers of the figure bodies in a set of 220 full-body paintings.

was estimated. Pose embedding method was also proposed to learn an embedding that places images of humans in similar poses nearby [42], [43]. It did not need to estimate body joint positions. The current research on pose retrieval generally indicates that the pose representation by particular location for each body part or particular angle between body parts is sensitive to the body part detection error. Soft coding of pose, which is to consider several likely alternative locations in constructing the pose representation, could improve the performance of pose retrieval.

III. Composition in Portrait Paintings

Portrait paintings collected for this study are chosen from artists of the 17-19 centuries such as: Thomas Gainsborough (1727–1788), Fransisco Goya (1746–1828), Edouard Manet (1832–1883), and Jean-Auguste-Dominique Ingres (1780–1867), and so on. These artists maintained a high dynamic range in their paintings which was broadly similar to that of photographs. Additionally, they painted a lot of portraits with an outdoor background. In total, 500 portrait paintings are collected to form the database. The database covers a variety of pose and environments. 280 of them are half body portrait paintings and 220 of them are full body portrait paintings. Based on the size of the face in the image, the half body portrait paintings are classified to small face half body portrait (191 paintings) and big face half body portrait (89 paintings, face length/image height ratio > 0.2). The paintings in the database are in height all larger than 500 pixels.

In portrait paintings, the pose of the figure and the placement of the face are commonly considered together [44], [45]. Three example paintings with figures in different poses are shown in Fig. 1. In the painting in Fig. 1(a), the body of the figure has a left rotation, the figure is centered at the vertical central line, but the face is shifted to the left of the image. In the painting in Fig. 1(b), the body of the figure has a right rotation and the face and the figure are biased to the right of the image. While the faces in the two paintings in Fig. 1(a) and (b) are all on the top of the one third line, the face in Fig. 1(c) is around the one third line. The placements of the faces in the three paintings are clearly different as the figures are in different poses. The centers of faces and centers of the bodies in 220 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. Faces are placed based on the pose of figures in the paintings.

IV. Proposed Composition Improvement Method

The observation of composition in portrait paintings provides the justification for the improvement of the composition of portrait photographs based on the pose of the figure. The framework of the proposed method is shown 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, the pose contains both the body pose and face pose. The face pose is comprised of the face tilt direction and the face rotation angle (i.e. face viewpoint). The face tilt direction can be estimated together with the body pose by the pose estimation method. Therefore, in the following content, pose refers to the body pose together with the face tilt direction. The face rotation angle is extracted separately as the face direction.

Given an input portrait photograph, the pose and face direction are extracted firstly. Apart from the pose and face direction, the location of the figure is also considered in the example painting selection. The location of the figure is defined by spaces around the figure. These space arrangement features are to ensure that the photograph can be cropped so that it has a similar space arrangement to that of the selected example painting.

Because the pose and the 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 $G = (V, E)$ can be 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 two types of nodes in the graph. One is the

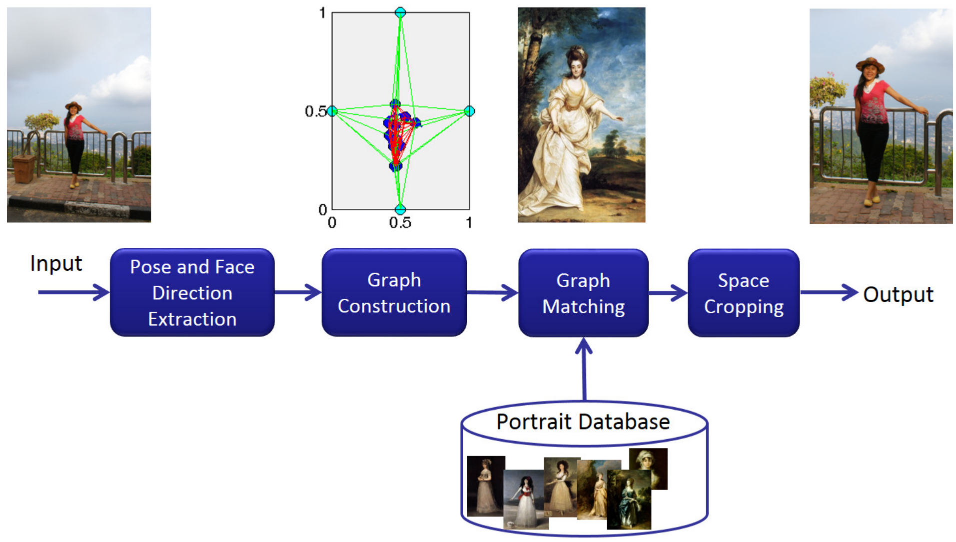


Fig. 2. The framework of the proposed method.

body node which represents the body part, and the other is the boundary node distributed on the boundaries of the image.

After constructing the graph, the example painting is selected by searching the painting database based on a similarity measure for graph matching. Then, space cropping is performed by using an optimizing function.

A. Pose extraction and correction

In our work, the full body pose and also half body pose all require to be extracted to guide the photograph composition improvement. The pose estimation needs to be conducted for both portrait photographs and paintings. By studying the current existing pose estimation methods including both traditional and deep learning methods [40], [37], [38], it indicates that the deep learning methods perform well for pose estimation of photographs. However, the pose estimation performance of these deep learning methods is still dramatically decreased for the pose extraction in paintings. The main reasons are the background color disturbance and there are a lot of body part occlusions due to the dressing style in paintings. In order to extract the accurate pose of the figure in the two situations above, we propose to use an interactive pose correction method. The user only needs to give several inputs to correct the wrongly detected body parts. It saves a lot of time comparing to completely manual pose generation.

Besides the pose estimation accuracy, the pose representation is another factor that affects the pose retrieval performance. The current research on pose retrieval generally indicates that the pose representation by particular location for each body part or particular angle between body parts is sensitive to the body part detection error. Soft coding of pose, which is to consider several likely alternative locations in constructing the pose representation, could improve the performance of the pose retrieval [41]. Therefore, we propose to use the probability distributions over possible part positions and orientations for the pose retrieval as in [40]. Based on the initially extracted pose, the user could interactively correct the

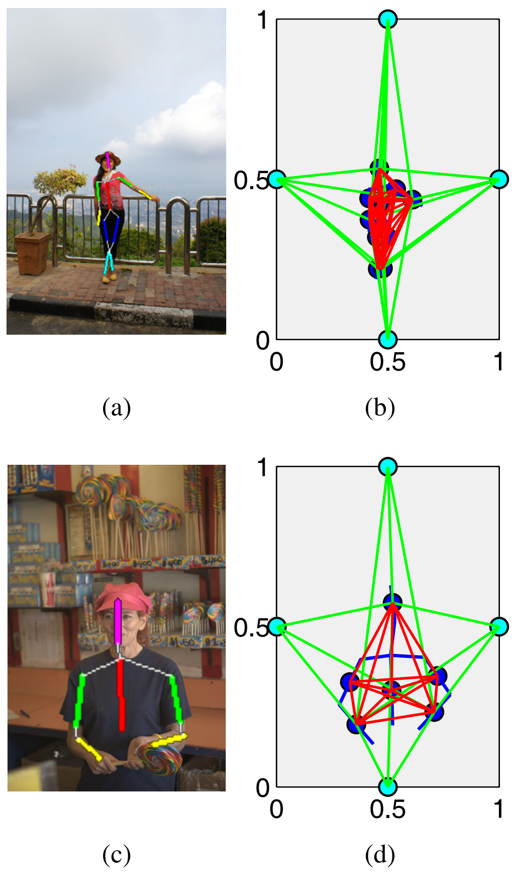


Fig. 3. (a) A full-body portrait photograph with extracted pose; (b) The graph for the photograph (a); (c) A half-body portrait photograph with extracted pose; (d) The graph for the photograph in (c). The nodes in dark blue are the body nodes, which express the body parts. The nodes in cyan are the boundary nodes. The blue lines are the skeleton of the body. The red lines are the edges linking the body nodes and the green lines are the edges linking the head, lower arm and lower leg nodes with the boundary nodes.

failure detected body parts. Then, the probability distributions over possible alternative part positions and orientations are calculated based on the pose correction constraints. They are used for the selection of the example portrait painting based on the similarity of the poses.

In our work, the pose of the figure is initially extracted by the pose estimation method proposed in [40] (other methods could also be used). For full-body pose, there are 10 body parts (including head, torso, left upper arm, right upper arm, left lower arm, right lower arm, left upper leg, right upper leg, left lower leg, right lower leg). For half-body pose, there are 6 body parts (including head, torso, left upper arm, right upper arm, left lower arm, right lower arm). The pose extraction is performed based on trained human pose estimators.

Control points are added in the initially extracted pose (see Fig. 4). The estimated pose achieved using the method in [40] includes the locations of the body part end points. They define the direction and location of each body part. The body part end points are activated as control points. The users could correct the extracted body part location and orientation by dragging the corresponding body part end control points. Then, the corrected body part location provides constraints on the location and direction for posterior estimation of a configuration of body parts L. Given an image I and $L = l_i$, where l_i is i-th body part, we have

$$P(L/I) \propto exp(\sum_{(i,j)\in E^b} \Psi(l_i, l_j) + \sum_i \Phi(l_i) + \sum_i \gamma(l_i) + \sum_i \eta(l_i)) \tag{1}$$

where $\Psi(l_i, l_j)$ is the pairwise potential which corresponds to a prior on the relative position of body parts. $\Phi(l_i)$ is the unary potential which corresponds to the local image evidence for a body part in a particular position. E^b is the pose model structure. $\gamma(l_i)$ is the location prior and $\eta(l_i)$ is the orientation prior. $\gamma(\cdot)$ gives uniform probability to a rectangular area centered at the location of the corrected body part and zero for other locations. $\eta(\cdot)$ gives maximum probability to the orientation θ_c of the corrected body part. The probability to other orientation values is reduced based on a Gaussian distribution which centered at θ_c. The estimated posterior marginal distribution $P(L)$, $L = l_i$ based on the constraints from the corrected body parts will be used for the graph matching in the example painting selection. Four examples of pose correction are shown in Fig. 5.

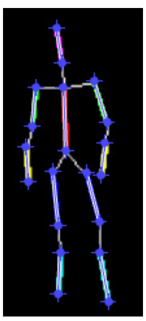


Fig. 4. Control points (in blue) for pose correction.

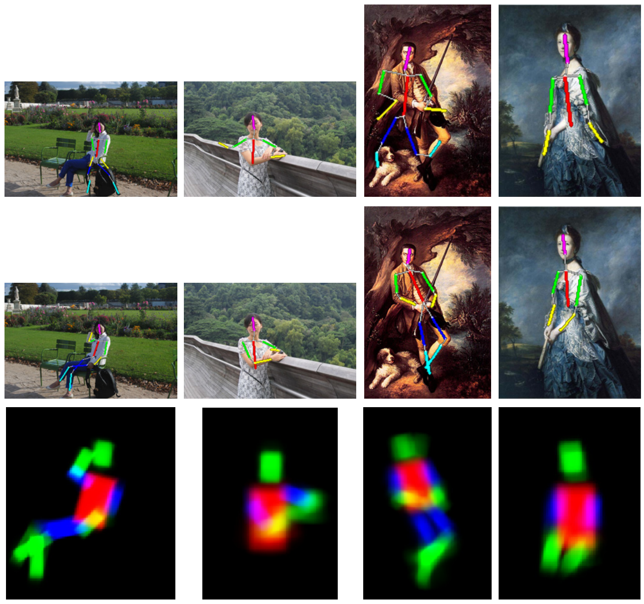


Fig. 5. Top:Original estimated pose; Middle:Corrected pose; Bottom: Estimated pose soft map based on the constraints, obtained by convolving rectangles representing body parts with their corresponding posterior.

B. Graph matching

After constructing the graph to describe the pose and spaces around the figure, the example painting selection is cast to find a painting that has a graph similar to the graph of the input photograph. Therefore, 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

$$\max_k S(G_I, G_k) \tag{2}$$

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

$$S(G_I, G_k) = \sum_i S_i^v(G_I, G_k) + \sum_j S_j^e(G_I, G_k) \tag{3}$$

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 the nodes and S^e is the similarity of the edges. The definition of similarities of edges and nodes is based on the definition of descriptors.

The descriptors of a body node v^b are the location and orientation of the corresponding body part. In the pose estimation, the posterior marginal distribution of the position (including location and orientation) of each body part is estimated in section IV-A. Based on the posterior marginal distribution, the position and orientation of each body node are calculated. For the body node v_i^b, the position descriptor is the posterior marginal distribution $P_{pos}^{v_i^b} = P(l_i = (x, y, \theta))$, where l_i is the body part i, (x, y) is the scale-normalized location, and θ is the orientation. The orientation descriptor of v_i^b is measured by $P_{ori}^{v_i^b} = \sum_{(x,y)} P(l_i = (x, y, \theta))$.

The descriptors of the edge e^b linking the body nodes are the relative location and orientation of the two corresponding body parts. The relative location descriptor of the edge e_j^b which links the body parts j_1 and j_2 is $P_{pos}^{e_j^b} = P(l_{j_1}^{xy} - l_{j_2}^{xy} = \delta)$, δ is the location distance value. The relative orientation descriptor of the edge e_j^b is $P_{ori}^{e_j^b} = P(r(l_{j_1}^{\theta} - l_{j_2}^{\theta}) = \rho)$, where $r(\cdot)$ is a circular difference operator and ρ is the orientation difference value. Details of the calculation of the descriptors can be found in [40].

Specifically, the face node is described by the face direction. The face direction is the face rotation angle along the yaw direction. It describes the face viewpoint. It is described by 13 quantized viewpoints at every 15° from 90° to -90° [46]. 90° is corresponding to the right side face viewpoint and -90° is corresponding to the left side face viewpoint. The face direction is detected together with the face location by fitting a linearly-parameterized, tree-structured face part model. The face part is defined at each facial landmark which is detected by HoG detectors. The shape of the tree model captures topological changes due to viewpoint. More details could be found in [46]. The descriptors of the four boundary nodes v^d are the boundary centers. The descriptor of the edge e^d that links the body node with the boundary is their distance in the x or y dimension.

As in [40], the combined Bhattacharyya similarity $s(a, b) = \sum_j \sqrt{a(j) \cdot b(j)}$ is used to calculate the similarities of the descriptors of body nodes v^b and edges e^b. The similarity of two face directions θ_1 and θ_2 is measured as

$$S_F(\theta_1, \theta_2) = e^{-(\frac{(\theta_1 - \theta_2)}{\sigma})^2} \tag{4}$$

where σ controls the sensitivity of the similarity to the difference of face directions.

The locations of the boundary nodes are the same for all the graphs, therefore it is not necessary to calculate the similarity of corresponding boundary nodes. The constraint on the distance of the boundary node with the body 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, the similarity of the q-th e^d edge from the input graph with the corresponding e^d edge from the k-th painting graph with descriptors d_I and d_k is

$$S_q^{e^d} = \begin{cases} 1 & \text{if } d_I - d_k \geq 0 \\ e^{-(\frac{d_I - d_k}{\sigma_d})^2} & \text{if } d_I - d_k < 0 \end{cases} \tag{5}$$

where σ_d is the standard variance.

After calculating the similarity of body nodes S^{v^b}, the similarity of face directions S_F, the similarity of edges linking body nodes S^{e^b}, and the similarity of edges linking body nodes with boundary nodes S^{e^d}, the final similarity of two graphs is the linear combination of these similarities, which is expressed as

$$S(G_I, G_k) = \sum_n S_n^{v^b} + S_F + \sum_p S_p^{e^b} + \sum_q S_q^{e^d} \tag{6}$$

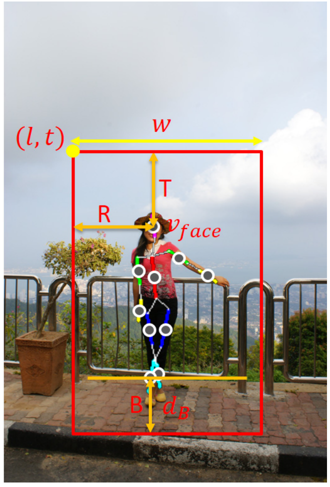


Fig. 6. Illustration of cropping window, body part location, and space around the figure. The cropping window is expressed by the left top corner coordinates (l, t) and width (w). The while cycled gray point is the center of each body part, representing the body part location. T is top, R is right, and B is bottom.

where $n = 1, ..., N^b$, N^b is the number of body nodes. $p = 1, ..., M^b$, M^b is the number of edges linking body nodes. $q = 1, ..., M^d$, M^d is the number of edges linking body nodes with boundary nodes. The painting that has the highest similarity score with the input photograph is selected as the example to guide the cropping of the input photograph.

C. 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 (red window in Fig. 6). 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 considering the location of each body part and the space around the figure. The energy function for the space cropping is defined as

$$E = E_p + \alpha E_s \tag{7}$$

The first term E_p is the pose constraint, and the second term E_s is the space constraint. Given the target location v_i for the body part i and the reference location v_i^r which is from the example painting, E_p is defined as

$$E_p = \frac{1}{N^b} \sum_i^{N^b} \|v_i - v_i^r\| \tag{8}$$

The body part location is defined as the center of each body part, shown as the white circled gray point in Fig.6.

The space around the figure includes space to the top, bottom, and side boundaries of the canvas. Due to the importance of the face in a portrait, the space to the top and side boundaries is represented by the distances of the face to the top and right side boundaries, which is also the face location v_{face} as illustrated in Fig.6. The space to the bottom boundary is represented by the shortest vertical distance of the body parts to the bottom boundary, expressed by d_B. Therefore, the space constraint E_s is measured by

$$E_s = \|v_{face} - v^r_{face}\| + |d_B - d^r_B| \tag{9}$$

where v^r_{face} is the reference face location and d^r_B is the reference distance to the bottom boundary. They are from the example painting. The pose constraint and the space constraint is 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) [47] method is used to seek the optimal solution by globally searching for the minimum candidate of (7).

V. Experiments

The algorithm was implemented in MATLAB on a PC with an Intel 2.67GHz processor and 4GB RAM. For a photograph of a pixel dimension of 1000×750, it takes around 0.95 seconds to obtain the optimization result using PSO. In the similarity measure of the face directions (4), the parameter is set as $\sigma = 45$ to reduce the similarity score to be small while the direction difference larger than 45°. In (5), σ_d is set empirically as 0.3. In the energy function (7) for space cropping, $\alpha = 0.3$ is applied for our experiments.

Based on the study of aspect ratios of portrait paintings, the aspect ratios (height/width) for full-body portraits and small face half-body portraits are defined as $a = 1.5$, and for big face half-body portraits, it is $a = \frac{4}{3}$. The test portrait photographs were collected from the MIT-Adobe FiveK Dataset [48] and our personal collections.

A. Performance

The cropping results of some photographs with figures in different poses are shown in Fig. 7. The pose of the figure in the selected example painting is similar to that of the corresponding input photograph. From composition improvement results in Fig. 7, we can see that the location of the figure in the result is similar to that in the corresponding example painting. This shows the effectiveness of the proposed method. Additionally, we can see that the locations of the figures in the composition improvement results are very different and they serves the requirement of their poses.

In order to analyze the affect of parameter α in (7) for space cropping, 40 test photographs with full body figures are cropped using the proposed method by setting α to different values. The space cropping results are also with full body figures. The matching errors of the 10 body part locations and space to bottom are measured to analyze the affect of α. The matching error of the body part is calculated as the Euclidean distance between the body part location in the result and that in the reference painting. The matching error of the space to bottom is calculated as the absolute difference between the value in the result and that in the reference painting. The width and height of each photograph and painting are normalized to 1 to measure the body part location and space to bottom. The average error of the 40 test photographs is summarized in Fig. 8. For values of α bigger than 0.5, the change of matching errors are very small. With the increasing of α from 0 to 0.5, the matching errors of the face and space to bottom are significantly reduced, in contrast, the matching errors of other 9 body parts are increased distinctly. It certainly reflects that the reducing of matching errors of face and space to bottom is at the cost of raising the matching errors of other body parts. To make the balance of matching errors, $\alpha = 0.3$ is recommended and it is used for our experiments.

The matching errors of body parts and space are generally small (Fig. 8), which are in error range of 10-60 pixels for an image in dimension of 1000×1000. This conveys the accuracy of our reference painting selection method. The lower arms and legs are more likely to have higher matching errors. The distribution deviations of them are also high, see Fig. 9. The reason is that the figures in selected reference paintings may have different placements of lower arms or legs from the input photograph.

B. Comparison with rule-based methods

To evaluate the effectiveness of the proposed composition improvement method, it is compared with two rule-based automatic composition improvement methods. These two rule-based methods are designed for photographs with a face. One method is the auto cropping based on templates in [17] (ACDP method). For an input photograph, a template is firstly selected based on the size of the face in the original photograph. Then, cropping is conducted based on the rule definition of the selected template. The other method is the aesthetic-based photo editing method proposed in [19] (ABPE method). The composition optimization is conducted by cropping based on the measurement of how well the face composition fits the rule of thirds, the visual balance, and the influence of the relative face sizes compared to the image. These two methods only generate one composition optimization result for each photograph.

For a full-body portrait photograph, the proposed method can crop it to a full-body portrait, a half-body portrait with small face, or a half-body portrait with big face. In the half-body portrait painting with big face, generally only the upper body with upper arms are visible. The extracted poses for lower arms are not meaningful. Therefore, in the selection of the example big face half-body portrait painting, the lower arms are not considered. For a small face half-body portrait photograph, the proposed method can also crop it to small face half-body and big face half-body portraits.

The composition improvement results for two of the test photographs are shown in Fig. 10. The result in Fig. 10(h1) and (h2) are produced by the ACDP method using the small face template. The selection of the template is based on the size of the face in the original photograph. In the ACDP method, the ROI is used to constrain the cropping to prevent the photograph from being cropped too aggressively. The ROI is detected by an attention model. The detected ROI

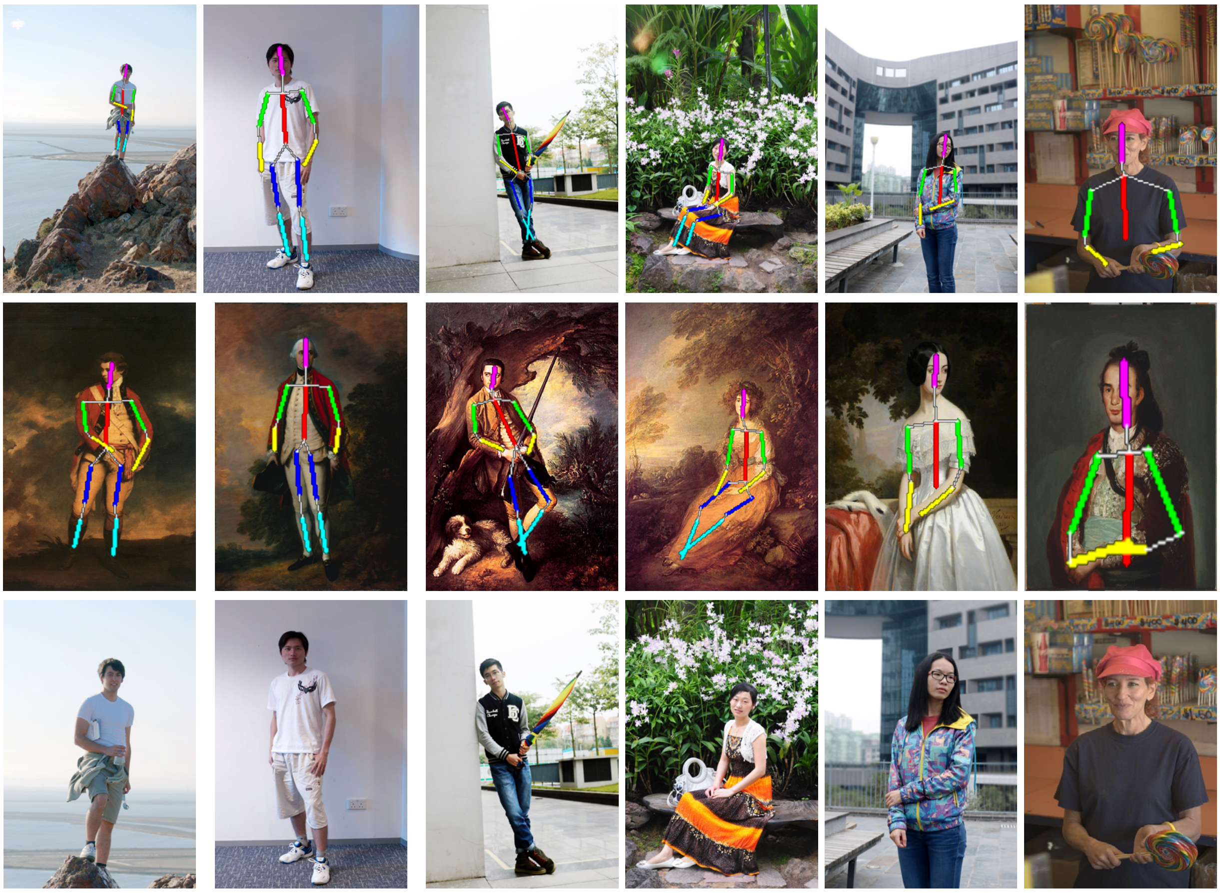


Fig. 7. Top:Photographs with extracted poses; Middle:Selected example paintings with extracted poses; Bottom:Cropping results based on selected example paintings.

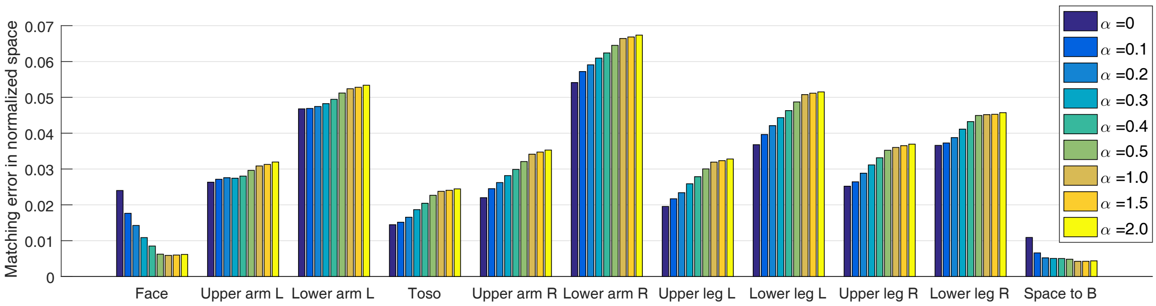


Fig. 8. Matching errors of the 10 body part locations and space to bottom (B). L is left, R is right. The matching error is measured by normalizing the width and height of each photograph to 1.

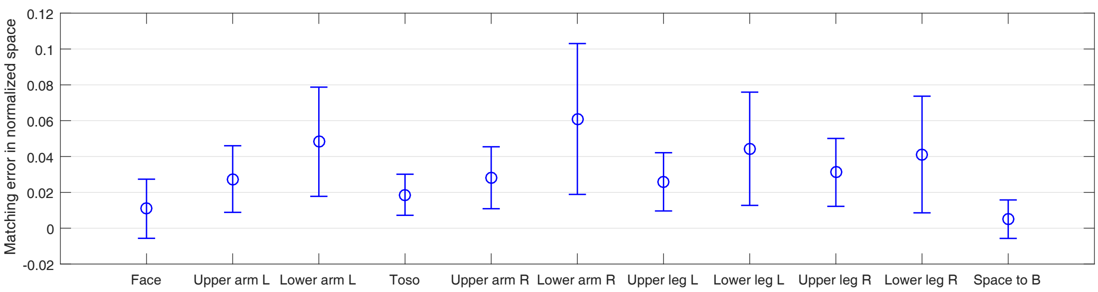


Fig. 9. The average matching errors of the 10 body part locations and space to bottom (B) while $\alpha = 0.3$. L is left, R is right. The matching error is measured by normalizing the width and height of each photograph to 1. The bars represent +/- one standard deviation.

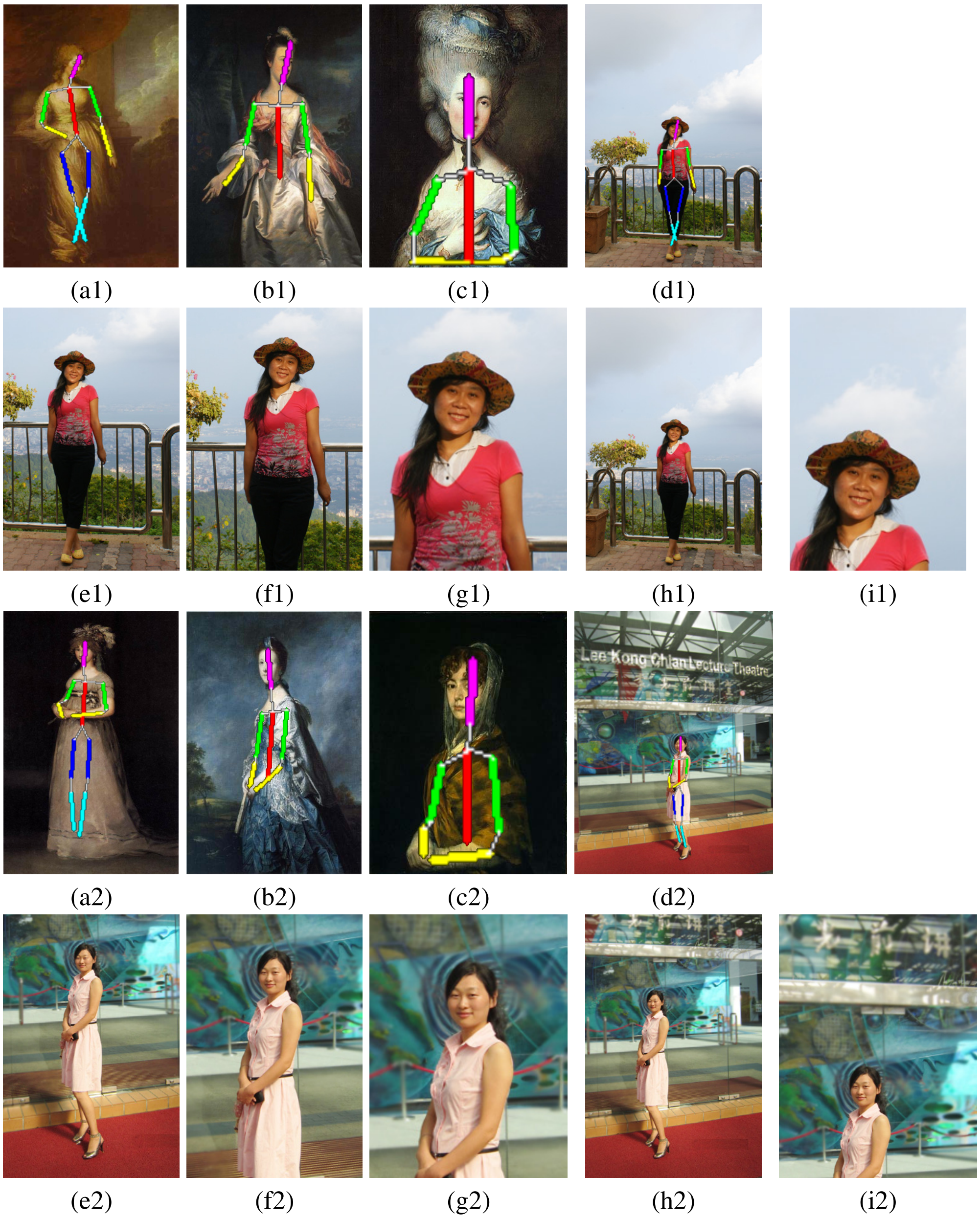


Fig. 10. (a1)-(c1) and (a2)-(c2) Example paintings with extracted pose; (d1) and (d2) Input photograph with extracted pose; (e1) and (e2) Space cropping result (full-body) by the proposed method based on example painting in (a1) and (a2); (f1) and (f2) Space cropping result (small face half-body) based on example painting in (b1) and (b2); (g1) and (g2) Space cropping result (big face half-body) based on example painting in (c1) and (c2); (h1) and (h2) The result by the ACDP method in [17]; (i1) and (i2) The result by the ABPE method in [19].

contains big regions of the background. Therefore, the space cropping effect is not distinct. The space cropping result in Fig. 10(h1) is almost the same with the original photograph. The ABPE method [19] only considers the face region as important information. Hence, only the face and upper body are visible while the other parts are all cropped out (see results in Fig. 10(i1) and (i2)). In the results in Fig. 10(i1) and (i2), the faces are placed close to the bottom one third points while leaving half of the space on the top for background. Although this placement of the face is better than its original placement, it is not reasonable while considering the visual appearance of the whole image. The results by the proposed method are more encouraging compared with those by the rule-based methods.

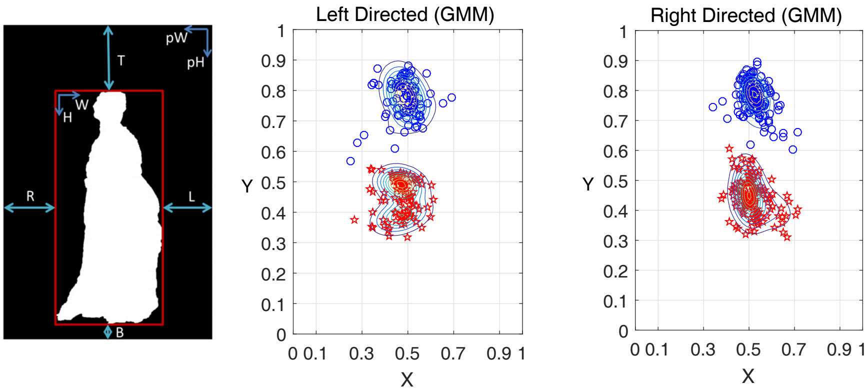
C. Comparison with learning-based methods

The proposed method is also compared to three of the learning-based image cropping methods [22], [26], [24]. The learning-based method proposed by Fang et al. [22] (we call it L-Fang method) trains a support vector machine (SVM) classifier to score cropping candidates. The SVM classifier is trained by using well-composed photographs considering spatial distribution of salient regions. Positive samples are obtained from these well-composed photographs and negative samples are created by generating random crops of these photographs. Finally, the score from the SVM classifier is fused with the boundary simplicity and content preservation scores to generate the final score for each crop candidate.

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

Fig. 11. Comparison with learning-based methods. (a) Input photographs; (b) Results by our method; (c) Results by the VFN [24]; (d) Results by the L-Huang method [26]; (e) Results by the L-Fang method [22].

In the method proposed by Huang et al. [26] (we call it L-Huang method), 13 features are extracted describing the visual representativeness and foreground recognizability to train a crop window classifier. The classifier is trained from a training dataset consisting of photographs before and after cropping by expert photographers. The cropping windows generated by the photographers are considered as positive samples. The negative samples are created by generating cropping windows which are significant different from those give by expert photographers. 2000 human cropped photographs with face were collected from the dataset released by [49] to train the L-Huang method. The method in [24] trains a view finding network (VFN), which is composed of a ConvNet augmented with a ranking layer. The VFN is trained using photographs by professional photographers from Flickr. The aspect ratio of the cropping results generated by VFN and L-Fang methods is dynamically decided based on the content of the input photographs. The aspect ratio of the cropping results generated by the L-Fang method is fixed to 3:4. Some cropping results generated by the three learning-based methods are shown in Fig. 11. The cropping effect achieved by the VFN method is not distinct for photographs with rich detailed background (e.g. the first and fourth photographs in Fig.11). The L-Huang method is designed for thumbnail generation. The results achieved by the L-Huang method may get some important body parts cropped out, such as the feet in the first photograph and the upper body in the fourth photograph in Fig. 11. The aspect ratio of some cropping results generated by the L-Fang method is not suitable for a portraiture.



(a) (b) (c)

Fig. 12. (a) The spaces around the figure: L (left space of the figure), R (right space of the figure), T (top space), B (bottom space). H and W are the height and width of the figure respectively, and pH and pW are the height and width of the image respectively. (b) GMM models of centers of faces and centers of bodies of right directed full-body portrait paintings. (c) GMM models of centers of faces and centers of left directed full-body portrait paintings.

D. Comparison with a method using statistical models

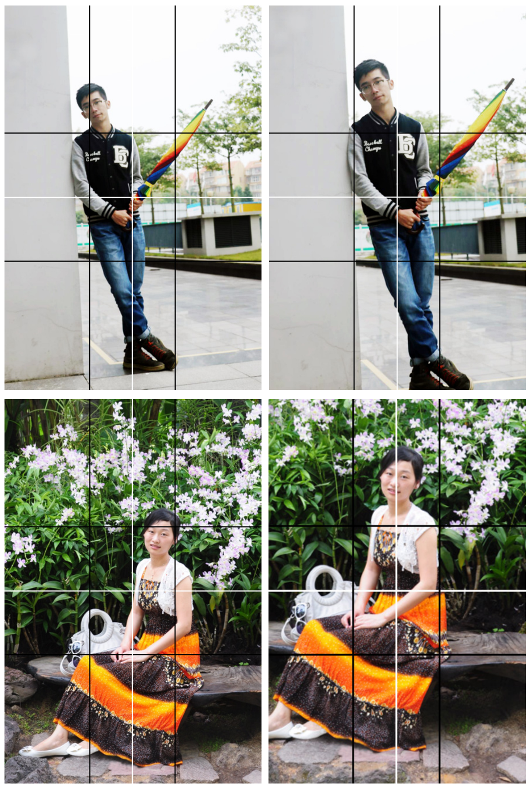
Besides the proposed composition improvement method based on the selected example painting, the method was also tested on cropping the portrait photograph based on the space arrangement from statistics. From portrait paintings, it was observed that the space was more likely to be larger on the left side of the figure if the head was facing left and more space was given on the right side of the figure if the head was facing right. Inspired by this, a study was conducted to explore the space arrangement in portrait paintings by dividing full-body portrait paintings, small face half-body portrait paintings and big face half-body portrait paintings into left directed and right directed classes based on the facing direction of the head. The spaces around the figure (see the Fig. 12(a)), the face location, the size of the face, and the center of the body were measured in each of the six classes. Then gaussian mixture models (GMM) were used to fit the distributions of these features. From the GMM models of locations of the faces (measured by the centers of the face windows) in the full-body portrait paintings shown in Fig. 12(b) and (c), we could observe that the center of the GMM model for the right directed class bias to the right of the central line, and it is opposite for the left directed class. The centers of the bodies are close to the central line.

Based on these statistical models, an energy function is formulated for each of the six classes. The energy function is

$$E = E_{TB} + E_{LR} + E_F + E_C + E_{FS} \qquad (10)$$

where, E_{TB} is the 2D GMM model of the top and bottom spaces, E_{LR} is the 2D GMM model of the left and right spaces, E_F is the GMM model of the face locations, E_C is the GMM model of the centers of bodies, and E_{FS} is the one dimensional GMM model of the face size. PSO method is used to maximize the energy function for finding an optimized cropping window.

This method based on statistical models creates baseline results (see the two examples in Fig. 13(b)). However, by using the statistical models, the face and the body are more likely to be placed close to the vertical central line. Differently, the proposed example-based method can produce results with faces and bodies placed in various locations based on the pose of the figure. Fig. 13 shows two results achieved using the proposed method compared to results achieved using the statistical method. The faces in the former are more biased from the central line than those of the latter. The face size is relatively larger in the results generated by using statistical models in Fig.13(b) than that in results generated by our



(a) (b)

Fig. 13. Comparison with results using statistical models. (a) Cropping results by the proposed method, and the original photograph and the reference painting are in Fig. 7. (b) Results using statistical models. The dark lines in the painting are the one-of-third lines and the white lines are the vertical and horizontal central lines.

method in Fig.13(a). Larger face may look more attractive at the first glance. However, the feet of the figures in the results generated by using statistical models are partly cut out. It results in that the figure is not intact for full body portrait and this cropping is either not a half body portrait. This is considered as a poor cropping in the artist view. The artists generally avoid cropping at close regions of any of the joints of the body which includes ankles and toes [50].

E. User study

In order to significantly evaluate the effectiveness and advantage of the proposed approach, a user study was conducted to compare composition improvements by our proposed method with those related methods. The user study was conducted in two groups. In the first group user study, our method was compared to ACDP [17] and ABPE [19] methods, and the method using statistical models. In the second group user study, our proposed method was compared to the three learning-based methods of L-Fang [22], L-Huang [26], and VFN [24] methods. 46 of the test photographs were randomly selected for this user study, including a variety of scenes. The participants did the user study using their own computers and monitors.

The results obtained by our proposed method and from one of the other methods were shown as a pair side by side. The left/right ordering of the images in each pair was randomly generated. At the starting page of the user study, a simple introduction to composition in portraiture was given. Then, for each image pair, 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?". The participants could choose "Left", "Right", or "They look the same" as a response. There were a total of 52 participants (33 male and 19 female) aged from 20 to 35 in the first group user study. 27 responses were collected in average for each image pair comparison. The result of the user study is summarized in Table I. It shows that our results are preferred by a significantly higher percentage of responses than those by other methods. The positive responses to our results are 6 times greater than those to ABPE method. The fewer responses to ABPE method is due to the fact that it only considers the face region as being important information and forces the face to the one third point without considering the other body parts. In some cases, the face is placed on the bottom one third point (as the example in Fig. 10 (i1) and (i2)). This results in the figure only taking a small space in the image while leaving a large space for the background. This comparison between our proposed method and ABPE method demonstrates that it is important to consider the face and body together in composition improvement. It is not surprising that our results are preferred more in comparison with results achieved using ACDP method, because ACDP method is easily disturbed by background objects. The method using statistical models is more likely to generate similar composition improvement results to our method than ACDP and ABPE methods. It could produce baseline results, however it is not as good as the proposed example-based method in the processing of images with diverse scenes.

TABLE I
USER STUDY RESULT FOR THE COMPARISON OF THE PROPOSED METHOD WITH THE RELATED METHODS.

First group	Our	vs Method	Same
Our vs ACDP	65.0%	24.9%	10.1%
Our vs ABPE	83.2%	12.9%	3.9%
Our vs Statistical Method	61.2%	23.3%	15.5%
Second group	Our	vs Method	Same
Our vs VFN	61.9%	26.1%	12.0%
Our vs L-Huang	57.2%	18.7%	24.1%
Our vs L-Fang	67.9%	21.7%	10.3%

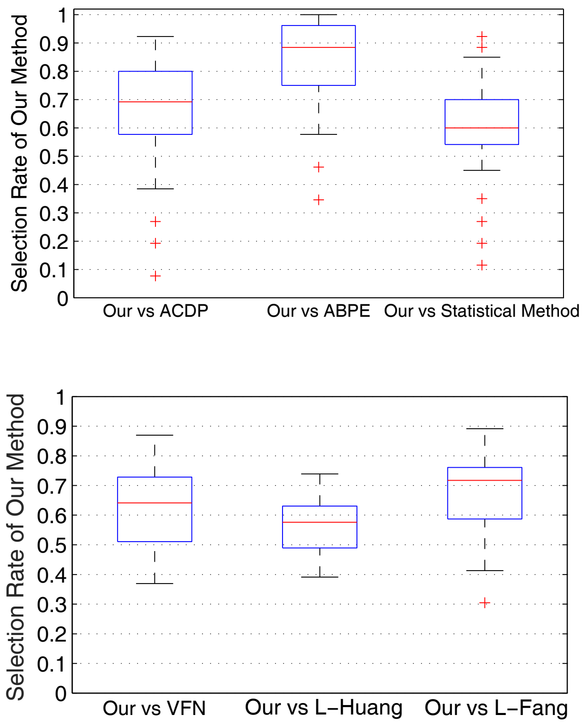


Fig. 14. Boxplot of the selection rates of our method by the participants. The central line in the box is the median of the distribution, and the edges of the box are the 25th and 75th percentiles

In the second group user study, there were 120 participants (69 male and 51 female) aged from 20 to 50. 40 responses were collected for each image pair comparison. The result of the user study in Table I shows that our results are also preferred by a significantly higher percentage of responses than those by the three learning-based methods. The positive responses to our results are around 2 times greater than those to the learning-based methods. The learning-based methods mainly consider the saliency region and face as important information in the composition without considering the other body parts. Additionally, in the classifier training, they could not get all the potential negative samples. These might result in unsuitable body part cropping and placing.

The box plot of the selection rates of all the participants for our method is shown in Fig.14. It shows that the majority of participants prefer the result generated by our method (selection rate is bigger than 50%). This conveys the coherence of human preference for our composition improvement results.

However, a few participants have a different view and their selection rates for our method may be less than 20%. This is not surprised as aesthetic preference is not innate and it is influenced by many social factors.

F. Discussion

The experiments and user study have demonstrated the effectiveness of our proposed approach. However, our approach has its limitations. The figure pose in the portrait paintings we used is very formal. For this reason, we could not always find a match in our database for some very informal photographs. Two failure matching cases are shown in Fig. 15. The figure poses in the selected examples are not well matching with those in the photographs. It results in that the space arrangement on the left and right of the figure is quite different from that in the reference painting. The hands of the figure in the first case and the feet in the second case all touch the boundary of the cropping images. In our future work, more portrait paintings will be collected. Additionally, the proposed method will be extended to use professional photographs as the examples to guide the composition improvement for input photographs.

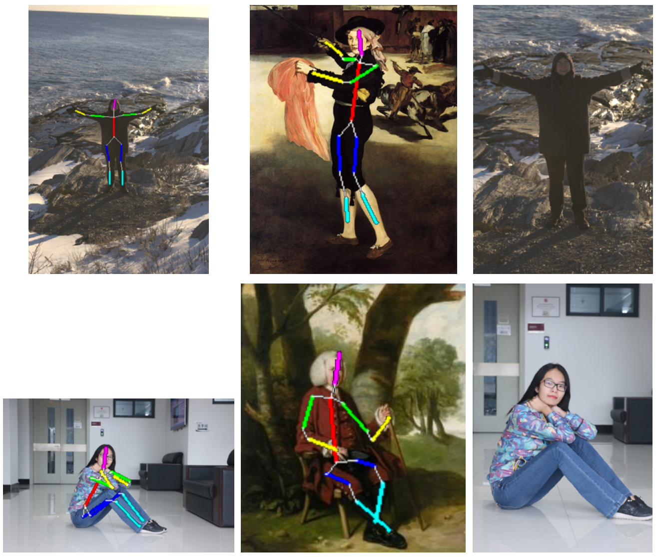

(a) (b) (c)

Fig. 15. Failure cases. (a) Photographs with pose; (b) Selected paintings with pose; (c) Space cropping results.

Our proposed method only considers the importance of figure and face in the composition improvement. The importance of background objects is not studied as in the context of portraiture nothing should compete with the face in terms of interest. Some important information in the background may be cropped out. For example, the mountain in the first image and the building in the fifth image of Fig. 7, and the title of the building in Fig. 10 (d2) are cropped out in our composition improvement results. The tile of the building in Fig. 10 (d2) is preserved partly in the cropping result by ACDP method (Fig. 10 (h2)). The ACDP method tries to preserve the important information of the image in cropping. However, in the comparison of our result in Fig. 10 (e2) to that by ACDP method in Fig. 10 (h2) in the user study, our result is preferred by 76% of participants. There may be important information in the background that serves to contextualise the portrait. For example, a holiday photograph where the figure is captured standing in front of a famous building or location. In our further study, we may explore the effect of background on aesthetics of portraits.

Some more experimental results by our proposed method are shown in Fig. 16.

VI. Conclusion

This paper proposes a method to improve the composition of a portrait photograph based on an example portrait painting. The example painting is selected based on the pose, the face direction of the figure, and the location of the figure. In order to avoid the influence of pose estimation accuracy on the composition improvement, a pose correction module with pose correction constrained posterior distribution calculation model is developed. A graph model is constructed for the example painting selection. Space cropping technique is used to improve the composition of the input photograph based on the figure pose in the selected example painting. The space cropping is formulated as an optimization problem. The experimental results and a user study show that the proposed method outperforms rule-based methods.

Acknowledgment

This work was supported in part by: (i) the National Natural Science Foundation of China (Grant No. 61602313, 61620106008, and 61602312); (ii) Shenzhen Commission of Scientific Research & Innovations under the Grant No. JCYJ20170302153632883, and JCYJ20160226191842793; (iii) Tencent "Rhinoceros Birds" - Scientific Research Foundation for Young Teachers of Shenzhen University; (iv) Research Foundation of Shenzhen University(2016051);(v)Startup Foundation for Advanced Talents, Shenzhen.

References

[1] A. Hess, *Borders Composition Digital Field Guide*. Wiley Publishing, 2010.
[2] W. H. Vanderburg, *Our Battle for the Human Spirit: Scientific Knowing, Technical Doing, and Daily Living*. University of Toronto Press, 2016.
[3] W. Blake, *Oil Portraits: Step by Step*. Courier Dover Publications, 1998.
[4] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1657–1664.
[5] H.-J. Lee, K.-S. Hong, H. Kang, , and S. Lee, "Photo aesthetics analysis via dcnn feature encoding," *ACM Transactions on Multimedia*, 2016.
[6] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 236–252, 2009.
[7] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proc. of International Conference on Multimedia*, 2010, pp. 271–280.
[8] S. S. Khan and D. Vogel, "Evaluating visual aesthetics in photographic portraiture," in *Proc. of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, 2012, pp. 55–62.
[9] S. Battiato, M. Moltisanti, F. Ravi, A. R. Bruna, and F. Naccari, "Aesthetic scoring of digital portraits for consumer applications," in *Proc. of SPIE, Digital Photography IX*, vol. 8660, 2013, pp. 7:1–7:10.

Fig. 16. More experimental results. First and fourth rows: Input photographs with pose, Second and fifth rows: Selected example portrait paintings with pose, Third and sixth rows: Cropping results using the proposed method.

[10] C. Li, A. C. Gallagher, A. C. Loui, and T. Chen, "Aesthetic quality assessment of consumer photos with faces," in *Proc. of International Conference on Image Processing*, 2010, pp. 3221–3224.

[11] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment," in *Proc. of International Conference on Computer Vision*, 2011, pp. 2206–2213.

[12] L. Huang, T. Xia, J. Wan, Y. Zhang, and S. Lin, "Personalized portraits ranking," in *Proc. of the 19th ACM International Conference on Multimedia*, 2011, pp. 1277–1280.

[13] B. Yan, K. Sun, and L. Liu, "Matching-area-based seam carving for video retargeting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 302–310, 2013.

[14] B. Yan, K. Li, X. Yang, and T. Hu, "Seam searching-based pixel fusion for image retargeting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 15–23, 2015.

[15] R. Hong, L. Zhang, and D. Tao, "Unified photo enhancement by discovering aesthetic communities from flickr," *IEEE Transactions on Image Processing*, vol. 25, pp. 1124–1135, 2016.

[16] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, vol. 29, no. 6, pp. 160:1–160:9, 2010.

[17] M. Zhang, L. Zhang, Y. Sun, L. Feng, and W. Ma, "Auto cropping for digital photographs," in *Proc. of the IEEE Conference on Multimedia and Expo*, 2005, pp. 438–441.

[18] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, pp. 469–478, 2010.

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

[20] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, and C. Chen, "Probabilistic graphlet transfer for photo cropping," *IEEE Transactions on Image Processing*, vol. 22, pp. 802–815, 2013.

[21] B. Ni, M. Xu, B. Cheng, M. Wang, S. Yan, and Q. Tian, "Learning to photograph: A compositional perspective," *IEEE Transactions on Multimedia*, vol. 15, pp. 1138–1151, 2013.

[22] C. Fang, Z. Lin, R. Mech, and X. Shen, "Automatic image cropping using visual composition,boundary simplicity and content preservation models," in *Proc. of ACM International Conference on Multimedia*, 2014, pp. 1105–1108.

[23] R. Hong, Y. Yang, M. Wang, and X.-S. Hua, "Learning visual semantic relationships for efficient visual retrieval," *IEEE Transactions on Big Data*, vol. 1, no. 4, pp. 152–161, 2016.

[24] Y.-L. Chen, J. Klopp, M. Sun, S.-Y. Chien, and K.-L. Ma, "Learning to compose with professional photographs on the web," *arXiv:1702.00503 [cs.CV]*, 2017.

[25] F. Farhat, M. M. Kamani, S. Mishra, and J. Z. Wang, "Intelligent Portrait Composition Assistance," in *Proc. of ACM Multimedia Conference*, pp. 17-25, 2017.

[26] J. Huang, H. Chen, B. Wang, and S. Lin, "Automatic thumbnail generation based on visual representativeness and foreground recognizability," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2015, pp. 253–261.

[27] Y. Wang, M. Song, D. Tao, Y. Rui, J. Bu, A. C. Tsoi, S. Zhuo, and P. Tan, "Where2stand: A human position recommendation system for souvenir photography," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, pp. 39:1–22, 2015.

[28] X. Zhang, K. L. Chan, and M. Constable, "Pose-based composition improvement for portrait photographs," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017.

[29] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 1, 2014, pp. 1653–1660.

[30] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Proc. of the International Conference on Neural Information Processing Systems (NIPS)*, vol. 1, 2014, pp. 1799–1807.

[31] I. Marras, P. Palasek, and I. Patras, "Deep globally constrained mrfs for human pose estimation," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2017, pp. 3466–3475.

[32] G. Ning, Z. Zhang, and Z. He, "Knowledge-guided deep fractal neural networks for human pose estimation," *IEEE Transactions on Multimedia*, vol. 1, pp. 1–14, 2017.

[33] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 1, 2016, pp. 4724–4732.

[34] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. of European Conference on Computer Vision(ECCV)*, vol. 1, 2016, pp. 483–499.

[35] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, "Multi-context attention for human pose estimation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 1, 2017, pp. 5669–5678.

[36] K. Sun, C. Lan, J. Xing, D. Liu, W. Zeng, and J. Wang, "Human pose estimation using global and local normalization," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2017, pp. 1–9.

[37] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *Proc. of European Conference on Computer Vision(ECCV)*, vol. 42, 2016, pp. 34–50.

[38] E. Insafutdinov, M. Andriluka, L. Pishchulin, S. Tang, E. Levinkov, B. Andres, and B. Schiele, "Arttrack: Articulated multi-person tracking in the wild," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 1, 2017, pp. 1293–1301.

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

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

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

[42] G. Mori, C. Pantofaru, N. Kothari, T. Leung, G. Toderici, A. Toshev, , and W. Yang, "Pose embeddings: A deep architecture for learning to match human poses," *arXiv:1507.00302v1*, pp. 1–9, 2015.

[43] N. Jammalamadakaa, A. Zissermanb, and J. C.V, "Human pose search using deep networks," *Image and Vision Computing*, vol. 59, pp. 31–43, 2017.

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

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

[46] X. Zhu and D. 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.

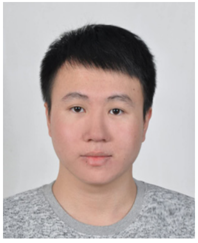
[47] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of the IEEE International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948.

[48] V. Bychkovsky, S. Paris, E. Chan, and F. 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.

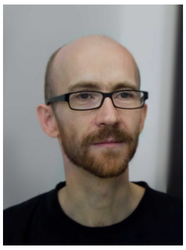
[49] S. A. Esmaeili, B. Singh, and L. S. Davis, "Fast-at: Fast automatic thumbnail generation using deep neural networks," in *arXiv:1612.04811v2*, vol. 1, 2017, pp. 1–9.

[50] E. Valind, *Portrait Photography: From Snapshots to Great Shots*. Peachpit Press, 2014.

Xiaoyan Zhang (S'10-M'16) received the Ph.D degrees from School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2014. She was a research fellow in Houston Methodist Research Institute, Houston, US, from 2013 to 2015. She is currently a lecturer in College of Computer Science & Software Engineering, Shenzhen University, China. Her current research interests include 2D image rendering, image enhancement, medical image processing, and 3D object modelling.



Zhuopeng Li received his B.S. degree from the School of Heilongjiang University, Heilongjiang, China in 2017. He is currently working toward the M.S. degree in the Department of computer science and technology, College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China. His current research interests include image processing, deep learning and reinforcement learning.



Martin Constable received the B.A. degree in Fine Art from the Camberwell College, London, England, in 1986, and an M.A. in Visual Art from Goldsmiths' College, University of London in 1990. He has taught painting, drawing, art history and studio practice at the Royal College, London, Goldsmiths' College and Ruskin, Oxford. He was an Assistant Professor in the School of Art Design and Media, Nanyang Technological University, Singapore from 2006 to 2015. He currently working in RMIT University, Vietnam.

As a practicing artist he was winner of the prestigious Discerning Eye award in 1991 and has widely exhibited in Europe and Asia where he was the founding member of the Grieve Perspective collaborative. He also collaborates with engineers in the field of computational aesthetics: specifically, how the aesthetic attributes of an artwork might be computationally defined.

Kap Luk Chan (S'88-M'90) received the Ph.D. degree in robot vision from the Imperial College of Science, Technology, and Medicine, University of London, London, U.K., in 1991. He was an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2001 to 2015. Since 2015, he has worked in industry as director of technology in several Singapore companies. He is the founder of Tolendata Singapore R&D Centre Private Limited. He is now the director of AI centre in Yau Lee Holdings Limited, Hong Kong. His current research interests include image analysis and computer vision, particularly in statistical image analysis, application of machine learning and computer vision in robotics.

Zhenhua Tang received his B.S. degree in Mathematics and Applied Mathematics, Hefei University of Technology, Anhui, China in 2015. He is currently working toward the M.S. degree in the Department of Computer Technology, College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China. His current research interests include pose estimation and deep learning.

Gaoyang Tang received her B.S. degree from the School of Beijing Institute of Technology, Zhuhai, China in 2015. She is currently working toward the M.S. degree in the Department of software engineering, College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China. Her current research interests include digital image processing and deep learning.