

Improving Photo Composition Elegantly: Considering Image Similarity During Composition Optimization

Y. W. Guo¹ †, M. Liu¹, T. T. Gu¹, and W. P. Wang²

¹National Key Lab for Novel Software Technology, Nanjing University

²Department of Computer Science, The University of Hong Kong

Abstract

Optimization of images with bad compositions has attracted increasing attention in recent years. Previous methods however seldomly consider image similarity when improving composition aesthetics. This may lead to significant content changes or bring large distortions, resulting in an unpleasant user experience. In this paper, we present a new algorithm for improving image composition aesthetics, while retaining faithful, as much as possible, to the original image content. Our method computes an improved image using a unified model of composition aesthetics and image similarity. The term of composition aesthetics obeys the rule of thirds and aims to enhance image composition. The similarity term in contrast penalizes image difference and distortion caused by composition adjustment. We use an edge-based measure of structure similarity which nearly coincides with human visual perception to compare the optimized image with the original one. We describe an effective scheme to generate the optimized image with the objective model. Our algorithm is able to produce the recomposed images with minimal visual distortions in an elegant and user controllable manner. We show the superiority of our algorithm by comparing our results with those by previous methods.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms

1. Introduction

With the continuous performance improvement of digital cameras, humans can capture high quality photographs without suffering from the traditional factors such as noises, low contrast, and blur that may degrade photo quality, more easily than before. However, image composition as a crucial aspect influencing visual aesthetics is often ignored by most amateur photographers. Taking a high quality photograph with a good composition generally needs professional photography knowledge. A simple, yet intuitive guideline is the rule of thirds which means that an image should be imaged as divided into nine equal parts by two equally-spaced horizontal lines and two equally-spaced vertical lines, and important compositional elements should be placed along these lines or their intersections [Pet04].

To evaluate composition aesthetics and optimize photo composition automatically, the pioneering work is given by Bhattacharya et al. [BSS10] which learns a support vector regression model for capturing aesthetics. Image quality is enhanced by recomposing user selected salient object onto the inpainted background or by using a visual weight balancing technique. Liu et al. [LCWCO10] develop a computational means for evaluating composition aesthetics according to the rule of thirds and other visual cues. A compound operator of crop-and-retarget is used to modify the composition and to produce a maximally-aesthetic image.

When improving photo aesthetics by using automatic composition optimization techniques, the user may expect to maintain consistent visual perception over the resulting image as to the original one he shot. To account for this, the recomposed image should faithfully represent the original visual appearance as much as possible, rather than just visually pleasing. Previous methods on photo optimization

† Corresponding author: ywguo@nju.edu.cn

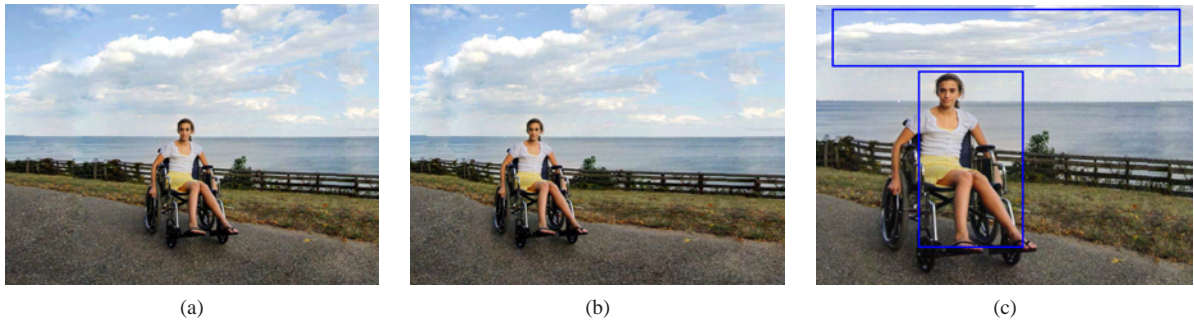


Figure 1: The image (b) with improved composition generated by our method for an input image (a). (c) shows the result of Liu et al. [LCWCO10]. The result exhibits obvious distortion in the cloud region. Moreover, the scale operation they adopt may cause blurring artifacts, see the girl sitting in the wheelchair.

however rarely consider this problem. As a result, the sudden and significant changes in image content may lead to an unpleasant user experience even though the optimal aesthetics is achieved. Another limitation of previous methods is that recomposing the prominent object to the optimal position suggested by aesthetic assessment directly, without any constraint on the result, may incur inevitable visual distortion, since inherently enhancing composition and reducing distortion conflict with each other, especially for those images with complex structures. As shown in Figure 1, the result (c) produced by the method in [LCWCO10] exhibits obvious distortion in the cloud region, and differs from the original image too much. This may be unacceptable to users.

In this paper, we present a new algorithm for improving the composition aesthetics of an input image while avoiding making significant changes to the visual appearance. Visual similarity between the optimized image and the original one is taken into account during the optimization of composition. The similarity is quantized by a measure of structural similarity called SSIM. We further incorporate a term of edge similarity into the similarity measure in order to reinforce the preservation of strong edges in images which are important visual cues. Our objective model combines the aesthetic measurement and the similarity. To compute the optimal image balancing composition aesthetics and visual similarity, we basically use seam carving to carve out a series of less noticeable seams and, correspondingly, to insert the same number of seams on the image. The optimal image is generated by searching the maximum of the objective model during this process. Our method can produce a good quality recomposed image in an elegant and user controllable manner. It is intuitive, easy-to-implement, and runs fast.

Our main contribution is a composition optimizing method which takes into account visual similarity between the optimized image and the original one. This allows us to produce the composition improved images which have minimal visual distortions, and retain faithful, as much as possible, to the original image content.

The remainder of the paper is organized as follows. The related work is introduced in Section 2. Section 3 describes our objective model that combines the composition aesthetics and image similarity. Section 4 shows how to compute the recomposed image maximizing the objective model. We conduct experiments and compare with previous methods in Section 5, and Section 6 concludes the whole paper.

2. Related Work

Photo quality assessment and enhancement, as an important aspect of computational photography, has attracted a large body of research. Those works on noise removal, brightness adjustment, and deblur are beyond the scope of this paper. We mainly review here the relevant methods on assessment and enhancement of composition aesthetics. Image retargeting, as an important means for improving composition, is briefly introduced.

2.1. Photo Composition Assessment and Enhancement

Ke et al. [KTJ06] propose a principled method to assess photo quality. High level semantic features are designed for measuring the perceptual differences between high quality professional photos and low quality snapshots. Different people, even for the professional photographers may have different aesthetic criteria in mind when taking and examining photographs. To bridge the gap between visual features and users' evaluation over quality, Bhattacharya et al. [BSS10] formulate photo quality evaluation as a machine learning problem in which the support vector regressors are used to learn the mappings from aesthetic features to visual attractiveness on composition. With the same features used to evaluate a given composition, the image with poor composition is enhanced by either relocating the segmented foreground onto painted background or balancing the visual weights of different image regions.

Liu et al. [LJW10] measure composition aesthetics based on the distributions of detected salient regions and prominent

lines. To modify image composition, a compound operator of crop-and-retarget is employed. The original parameter space is 6D, and the solution is found by particle swarm optimization in a reduced search space with some constraints. Cropping is used in this method. As concluded in [BSS10], this has two-fold problem. First, cropping reduces the size of image frame and can alter its aspect ratio. Second, cropping can lead to the loss of valuable image information in background which may be important to appreciate the images. Furthermore, re-scaling after cropping will inevitably blur the salient subjects without super-resolution. We in contrast prohibit the use of cropping and prevent the salient regions from size changes. The features on scene composition are characterized by analyzing and quantifying the locations and orientations of prominent lines in images [LWT11].

Some methods focus on improving other aspects on image aesthetics and achieve interesting results. The method in [COSG*06] enhances the harmony among the colors of a given photograph. Leyvand et al. [LCODL08] enhance the aesthetic appeal of human faces using a data-driven approach, while Zhou et al. [ZFL*10] improve the shapes of human bodies in images through a model-driven approach.

Recent research demonstrated that novel images can be generated by putting together components from different images according to a certain composition rules. In [CCT*09], given a freehand sketch annotated with text labels, a realistic picture is synthesized by seamlessly combining semantic components originating from different Internet photographs. In [HZZ11], Huang et al. present an approach for creating the so called Arcimboldo-like collage, which represents an image composed of multiple thematically-related cutouts from the filtered Internet images.

2.2. Image Retargeting

Composition optimization methods often use the image retargeting techniques to modify the positions of salient objects in resulting images. Retargeting refers to the process of adapting images to target screens of Cellular phones and PDAS which often have different resolutions and aspect ratios than the input.

Cropping is an efficient operation which cuts out important image regions for display [SAD*06, SLBJ03, NOSS09]. Salient image regions are first detected by saliency detection methods [IK01, CZM*11]. After cropping the surrounding area, important regions are reserved for display. Seam carving achieves remarkable image resizing results by iteratively carving out less noticeable seams. Recent methods achieve focus with context retargeting by fisheye transform based [LG05] or mesh-guided nonuniform content warping [WTSL08, GLS*09, ZCHM09]. An interesting work is recently given in [LJW10] which improves composition aesthetics during retargeting by using a mesh-based warping scheme. This technique may suffer from visual distortion in

some results. In our method, composition enhancement is coupled with image similarity, relieving perceivable artifacts of the resulting images.

Our work is also inspired by the retargeting techniques [RGWZ08, SCS108, BSFG09, RSA09] that consider similarity between the retargeted image and the input one as well. In [SCS108], image retargeting is framed as a maximization of bidirectional similarity between small patches of the original and output images. The same similarity measure is accelerated by Barnes et al. [BSFG09] via random sampling and content coherence based match propagation. Rubinstein et al. [RSA09] combine different retargeting operators in an optimal manner. A similarity measure termed bi-directional warping is used with dynamic programming to find an optimal path in the resizing space. Different from their methods, our goal is to balance the influence of composition improvements and the variation of image content, rather than image retargeting. The objective is therefore quite different in essence.

3. Our Objective Model

We believe that a visually pleasing image with improved composition should satisfy two properties. On the one hand, its composition is optimized based on a certain rules of photo aesthetics. On the other hand, the optimized image should contain as much as possible information from the original image and bring as few as possible visual artifacts. To meet the above requirements, our objective model unifies image composition aesthetics and the similarity between the resulting image and the original one into a single formulation,

$$E(\mathbf{I}_r) = \lambda E_e(\mathbf{I}_r) + (1 - \lambda) \cdot E_s(\mathbf{I}_r, \mathbf{I}_o), \quad (1)$$

where E_e represents the composition aesthetics of the optimized result \mathbf{I}_r . E_s denotes the similarity between \mathbf{I}_r and the original input \mathbf{I}_o of size $w \times h$. $\lambda \in [0, 1]$ is a parameter that is used to control the influence of the two terms. Our aim is to solve \mathbf{I}_r that maximizes the above function. Obviously, the bigger λ is, the better composition the resulting image will have. Otherwise, a small λ will make the result resemble the input image more closely. In implementation, we set it to 0.5 for balancing the influences of composition aesthetics and image similarity. The optimized image which achieves the maximum of the above energy function is the good quality image that have balanced aesthetics and similarity compared with the original image.

3.1. The Composition Aesthetics

We basically use rule of thirds (ROT) to evaluate the composition aesthetics for images with distinct foreground objects. Such images are very popular in personal photo collections, for example the photos of family members, our friends, pets, and interesting objects like flowers. Detection of salient regions is crucial for computing the composition aesthetics, as



Figure 2: Some photos taken by professional photographers.

well later optimization. We exploit the salient region detection method developed in [CZM*11]. Furthermore, we employ the Viola-Jones face detector to detect human faces. In case that saliency detection and face detector fail to extract the salient regions, the user is allowed to draw outlines of foreground subjects.

With the detected salient regions, we calculate composition aesthetics considering two aspects. The first aspect, also widely exploited by previous work, is the distance from the center of interest to the four points, also called power points, where lines of ROT intersect. The second is whether or not the salient objects are placed along the ROT lines. This also obeys the ROT guideline. For an given image \mathbf{I} , we then model $E_e(\mathbf{I})$ as,

$$E_e(\mathbf{I}) = 1/3E_{sec}(\mathbf{I}) + 2/3E_{lin}(\mathbf{I}), \quad (2)$$

where E_{sec} and E_{lin} account for the above two aspects separately. It is reasonable to emphasize E_{lin} by setting it a big coefficient. Imagine that when a professional photographer takes a photo with a prominent subject, he often puts it along with a vertical line of ROT, rather than restricting its center onto one power point rigidly. Some examples are shown in Figure 2.

For E_{sec} , we sum up the score of each region weighted by region area,

$$E_{sec}(\mathbf{I}) = \frac{1}{\sum_i S_i} \sum_i S_i \cdot \cos \left(\left(\frac{|p_{ix} - p_{ix}^s|}{w/3} + \frac{|p_{iy} - p_{iy}^s|}{h/3} \right) \cdot \frac{\pi}{2} \right), \quad (3)$$

where S_i is the area of a salient region R_i . p_i and p_i^s are the center of mass of R_i , and the closest power point to p_i , respectively. With the above formula, E_{sec} for the image with a single subject whose center of mass lies on one of the four power points is set to 1.

To compute E_{lin} , an optimal solution is to extract the medial axis for each salient region and compute the distance to the nearest lines used in ROT guideline. However, for most

images with a prominent subject for example a person or a high building, the medial axis is nearly a vertical line segment. We hereby use a vertical axis as a substitute for the medial axis. Such an axis can be computed easily by finding an axis that divides the salient object into two parts of the same area. E_{lin} is then calculated as,

$$E_{lin}(\mathbf{I}) = \frac{1}{\sum_i S_i} \sum_i S_i \cdot \cos \left(\frac{2dis(L_i, L^s)}{w/3} \cdot \frac{\pi}{2} \right), \quad (4)$$

where L_i and L^s are the vertical axis of R_i , and the nearest vertical lines of ROT to L_i separately. $dis(\cdot)$ is the Euclidean distance.

To ease exposition, we only explore the effect of salient regions following the guideline of rule of thirds. Region area is taken as coefficients in the above formulations, emphasizing the influence of large salient objects. For those images like landscapes or seascapes that lack a distinct foreground objects, it is intuitive to compute the composition aesthetics by detecting the prominent lines and computing the score using E_{lin} . It is worth noting that previous techniques have used a learnt support vector regression model [BSS10] or a computational means [LCWCO10] to capture image aesthetics by taking more aesthetic perspectives into account. Such models can be seamlessly integrated into our framework for evaluating composition aesthetics.

3.2. Image Similarity

To improve composition aesthetics, retargeting techniques are often employed to adjust the positions of distinct foreground objects. During this process, images are subject to visual distortions, especially for those with complex background structures. In order to control such distortions within acceptable tolerance, similarity measure should be used to quantify the visual difference between the optimized image and the original one. Traditional quality metrics such as mean squared error (MSE), although simple to calculate, are not very well matched to the perceived visual quality. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, a measure of structural similarity, called SSIM, that compares local patterns of pixel intensities that have been normalized for luminance and contrast is developed in [WBSS04]. Experiments on several publicly available subject-rated image databases show that SSIM values exhibit much better consistency with the qualitative visual appearance. We therefore basically adopt SSIM to measure the similarity between the improved image and the input.

SSIM is defined as,

$$SSIM(\mathbf{I}_r, \mathbf{I}_o) = [l(\mathbf{I}_r, \mathbf{I}_o)]^\alpha \cdot [c(\mathbf{I}_r, \mathbf{I}_o)]^\beta \cdot [s(\mathbf{I}_r, \mathbf{I}_o)]^\gamma, \quad (5)$$

where $l(\cdot)$, $c(\cdot)$, and $s(\cdot)$ compare the luminance, contrast, and structures between \mathbf{I}_r and \mathbf{I}_o , respectively. α , β , and γ are parameters used to control relative importance of the three components. In order to simplify the expression, they can be

uniformly set to 1 [WBSS04]. Let $\mu_{\mathbf{I}_r}$ and $\mu_{\mathbf{I}_o}$ denote mean intensity of \mathbf{I}_r and \mathbf{I}_o separately. $\sigma_{\mathbf{I}_r}$ and $\sigma_{\mathbf{I}_o}$ are standard deviations. $\sigma_{\mathbf{I}_r\mathbf{I}_o}$ represents the covariance of image vectors of \mathbf{I}_r and \mathbf{I}_o . $l(\cdot)$, $c(\cdot)$, and $s(\cdot)$ are expressed as,

$$l(\mathbf{I}_r, \mathbf{I}_o) = \frac{2\mu_{\mathbf{I}_r}\mu_{\mathbf{I}_o} + C_1}{\mu_{\mathbf{I}_r}^2 + \mu_{\mathbf{I}_o}^2 + C_1}, \quad (6)$$

$$c(\mathbf{I}_r, \mathbf{I}_o) = \frac{2\sigma_{\mathbf{I}_r}\sigma_{\mathbf{I}_o} + C_2}{\sigma_{\mathbf{I}_r}^2 + \sigma_{\mathbf{I}_o}^2 + C_2}, \quad (7)$$

$$s(\mathbf{I}_r, \mathbf{I}_o) = \frac{\sigma_{\mathbf{I}_r\mathbf{I}_o} + C_3}{\sigma_{\mathbf{I}_r}\sigma_{\mathbf{I}_o} + C_3}, \quad (8)$$

where C_1 , C_2 , and C_3 are constants for avoiding computational instability. As suggested in [WBSS04], C_1 and C_2 are set to $(K_1L)^2$ and $(K_2L)^2$ in which L is set to 255 for 8-bit grayscale images, and K_1 and K_2 can be set as 0.01 and 0.03 separately in implementation. C_3 is usually set to $C_2/2$ in practice.

The essence of SSIM in contrast to tradition metrics is to compare the structures of two images directly. Studies of cognitive psychology show that human visual perception is very sensitive to the strong edges in natural images. Since re-targeting techniques such as seam carving [AS07] or mesh-guided warping [WTSL08, GLS*09] may easily destroy or deform important edges, it is vital to reinforce edge similarity in the similarity measure. To account for this, we adopt a measure of edge similarity that compares strong edges and use it as a compliment to structure comparison in SSIM.

Sobel operators with a horizontal edge mask and a vertical one are first applied to the given image. This yields two edge maps by exploiting which we can easily compute a gradient magnitude and an orientation for each edge pixel. An edge orientation histogram with 8 bins in $0^\circ - 180^\circ$ can thus be built, and we use it to compute edge similarity as follows,

$$e(\mathbf{I}_r, \mathbf{I}_o) = \frac{\sigma_{\mathbf{H}_r\mathbf{H}_o} + C'_3}{\sigma_{\mathbf{H}_r}\sigma_{\mathbf{H}_o} + C'_3}, \quad (9)$$

where $\sigma_{\mathbf{H}_r}$ and $\sigma_{\mathbf{H}_o}$ denote standard deviations of the histogram vectors of \mathbf{I}_r and \mathbf{I}_o , separately. $\sigma_{\mathbf{H}_r\mathbf{H}_o}$ is the covariance of two histograms. Note that, the histogram vector is normalized with respect to image area. $C'_3 \ll \sigma_{\mathbf{H}_r\mathbf{H}_o}$, $\sigma_{\mathbf{H}_r}\sigma_{\mathbf{H}_o}$ is still a constant for ensuring computational stability. We set C'_3 to 0.0001 in our experiments.

Integrating the edge similarity into SSIM, image similarity between \mathbf{I}_r and \mathbf{I}_o is finally calculated by

$$E_s(\mathbf{I}_r, \mathbf{I}_o) = l(\mathbf{I}_r, \mathbf{I}_o) \cdot c(\mathbf{I}_r, \mathbf{I}_o) \cdot \left(\frac{s(\mathbf{I}_r, \mathbf{I}_o) + e(\mathbf{I}_r, \mathbf{I}_o)}{2} \right). \quad (10)$$

4. Optimization

The resulting image is the image maximizing the objective function $E(\mathbf{I})$,

$$\mathbf{I}_r = \arg \max_{\mathbf{I}} E(\mathbf{I}). \quad (11)$$

Previous techniques on photo composition enhancement typically transform the images by image re-targeting coupled with the cropping operation. We take seam carving [AS07] as the basic operation for improving the composition aesthetics. In addition, we assume here two constraints for meeting users' requirements in practice. First, image sizes and aspect ratios should not be altered. Cropping is not allowed in this sense as cropping can lead to the loss of valuable image information, even though it can prescribe a straightforward solution to optimal re-composition [BSS10]. Second, salient regions should be free from zooming in or out since zooming in directly inevitably blurs the subject, while zooming out reduces resolution. With the above constraints, we use a heuristic algorithm to find the solution corresponding to the optimal image.

Seam carving changes image size and aspect ratio by carving out a series of less noticeable seams. A seam is an optimal path of pixels from top to bottom, or left to right defined in terms of local energy. Such a seam can be found using dynamic programming. Note that, in our application to preserve completely the salient objects, local energy defined on salient regions is valued with the maximal value in the energy function of seam carving.

We observe that by iteratively carving out seams on one side of a salient object and inserting the same number of seams on the other side simultaneously, image composition can be modified without changing image size. The key problem is thus to determine k , the number of seams to be removed, together with the k seams to be removed and k new seams to be inserted such that the resulting image \mathbf{I} maximizes $E(\mathbf{I})$. We denote $\{s_i^d\}_{i=1}^R$ and $\{s_j^a\}_{j=1}^R$ as the seams as the candidates for removal and insertion. Each s_i^d or s_j^a is labelled as 1 if it is selected, and 0 not selected. The optimization is then formulated as a labelling problem which can be solved by 0-1 mixed integer programming. Global optimization on the parameter space is still computationally expensive and may get stuck in local optima easily. We hereby develop an efficient heuristic algorithm which finds the solution in two steps: determining optimal positions of the foreground subjects and inserting and removing a certain number of seams.

Determination of optimal positions. Given the original image \mathbf{I}_o , the closest power point and vertical line used in rule of thirds to each salient object are first computed. Since the aesthetic term E_e has an analytical expression, we can easily determine the target location each subject should move towards by maximizing it.

Insertion and removal of seams. With the optimal loca-



Figure 3: Insertion and removal of seams. (a) For the image with a monkey as the foreground, the monkey will move right one pixel if a seam (yellow) is inserted on its left. The red seam is removed for maintaining image size. (b) For the image with two separate subjects, seams inserted and removed for adjusting the location of one subject should not sacrifice composition of the other. The left subject will move right one pixel, and meanwhile, the right will move left one pixel if the yellow seams are inserted and red ones are removed.

tions, composition aesthetics is enhanced by inserting and removing a certain number of horizontal or (and) vertical seams. Without loss of generality, we illustrate the process by taking horizontal movement of salient objects as an example as shown in Figure 3. That is to say, vertical seams are inserted and removed. Most photographs with distinct foreground subjects have at most two subjects. For the images with an individual subject, it will move toward the target location by simply removing seams on the side of target location and inserting seams on the opposite side. It is however a little bit of trouble for the images with two separate subjects. For ease of exposition, we call the side of target location the positive side, and opposite side the negative side. To avoid conflict, seams inserted and removed for adjusting the location of one object should not sacrifice the composition of the other object.

We basically adopt the seams suggested by seam carving. At the beginning, the horizontal and vertical distances between positions of subjects and their optimal positions are computed. The maximum number of seams to be removed and inserted in each direction is then determined. Furthermore, the seams are generated by dynamic programming on the original image. By successively carving out a series of vertical or horizontal seams on the positive side and inserting new seams on negative side, $E(\mathbf{I})$ will increase at the beginning as composition aesthetics is enhanced, although image similarity is reduced. Nevertheless, when a certain amount of seams are removed, $E(\mathbf{I})$ reaches the maximum and removing more seams will make the image differ from \mathbf{I}_0 too much. The image version corresponding to the maximum of $E(\mathbf{I})$ is the resulting image.

Computation cost of the above algorithm is mainly consumed by the process of computing seams to be removed and inserted. Fortunately, we only need to compute them once on the original image. It runs very fast. In addition, the process

can be further accelerated by removing and inserting several seams, rather than only one seam in each step.

An advantage of the above heuristic approach that works by successively carving out and inserting a series of seams is that the process the image changes gradually is open to the user. As the seams removed and inserted can be saved with very little extra memory, the user can backtrack to the intermediate results without re-executing the whole algorithm once he finds that the final result differs from his input too much.

5. Experiments

We experimented with our algorithm on a variety of images. Some representative results are shown in Figures 4 and 7.

Figure 4 shows that image energies E_e , E_s , and E change with the number of seams inserted and removed. E_e in both rows will achieve the maximum if the foreground subjects are re-composed onto the optimal positions suggested by the aesthetics rules. However, under the control of similarity term E_s in objective model, E achieves the maximum in front of E_e with the increasing number of inserted and removed seams. That is to say, foreground subjects in the images are moved to the positions which are close to, but not exact, the optimal positions determined by aesthetic optimization.

Similar cases are shown in Figure 7 which demonstrates more challenging examples. The images have different resolutions ranging from 640×480 , 878×652 to 1613×1024 . For the images with simple background such as the 6th image in the 1st column and the 5th one in the 3rd column, the interest objects are moved to the optimal aesthetic positions after composition optimization. However for most photos with complex background structures, the foreground objects are re-located to the new positions near the optimal positions under the control of the term of image similarity. Our algorithm produces the aesthetically improved images, without noticeable visual distortions.

To demonstrate the effectiveness of our algorithm, we also compare against the results of previous representative methods as shown in Figures 1, 5, and 6. As shown in Figure 5, the results (b) produced by [LJW10] may present obvious distortions in background. See the regions surrounded by the blue rectangles. In contrast, our results (c) avoid such distortions and look visually pleasing. Visual dissimilarity between the results and inputs is penalized by the similarity term in our objective model. The comparison in Figure 6 shows that our result is comparable to the result of [BSS10]. The method of [BSS10] relies on user-guided foreground segmentation and background inpainting for recomposing the object onto repainted background. The former however is difficult for the images whose foreground and background share similar color appearances. The latter is challenging for

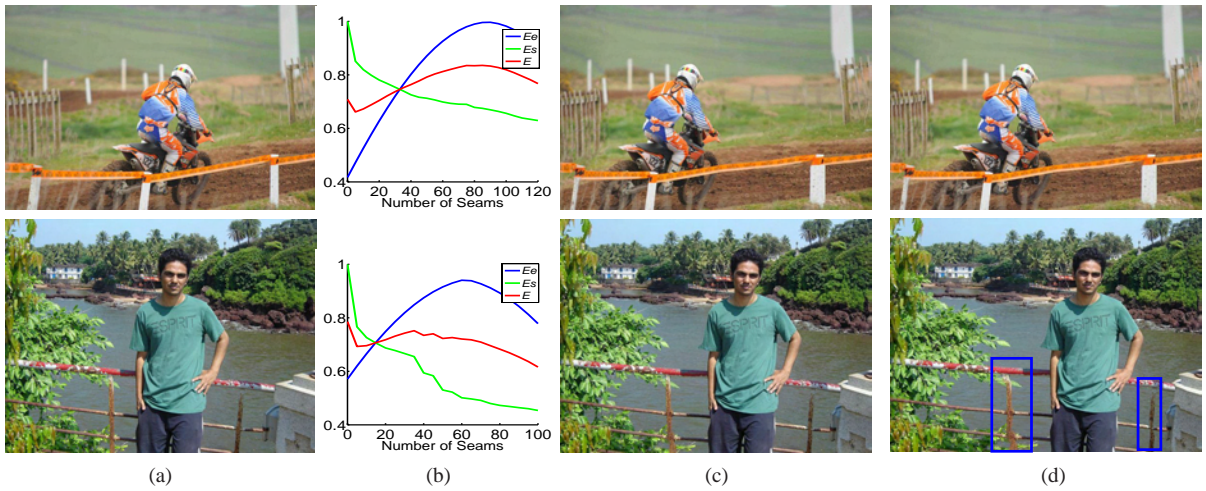


Figure 4: The resulting images (c) corresponding to the maximum of E shown in the red curves. (a) The inputs. (b) Variances of E_e (blue), E_s (green), and E (red) with the number of seams inserted (removed). (d) The resulting images by maximizing E_e only. For some images with complex background structures, visual distortions will be introduced if the input image is optimized with respect to composition aesthetics only. An example is shown in the second row, and the resulting image of (d) exhibits obvious distortions in the background area.

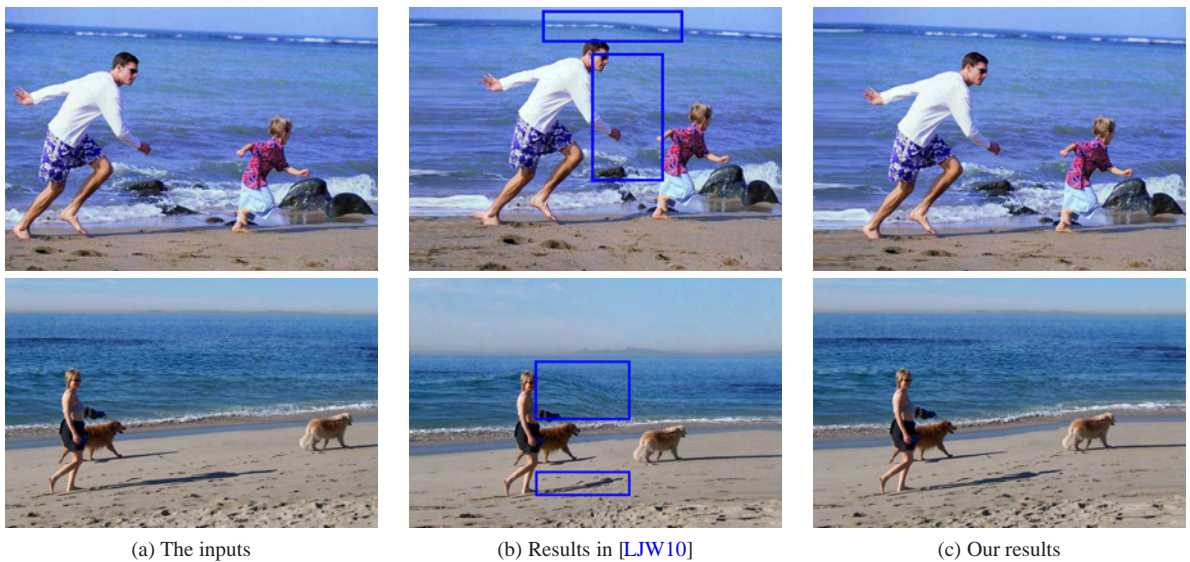


Figure 5: Comparison with the method by Liu et al. [LJW10]. (a) The input images published in [LJW10]. (b) Their method might suffer from noticeable distortions in the results. The distortions in water regions in both examples and the shadow in the bottom example are remarkable. (c) Our results.

those images with complex background structures. Furthermore, such scheme generally cannot handle the case where foreground object is occluded by the background region. An example is the image in the 1st row of Figure 4 where motorcycle tires are occluded by the fence. Our algorithm does not need to segment the foreground object and works well for such images.

Computational complexity of our algorithm mainly de-

pends on image size and the location of interest object, and the major computation is spent on computing seams using dynamic programming and measuring SSIM between the input image and the image series resulting from carving out and inserting seams. We use a fast multi-thread CPU implementation of seam carving, and also implement the fully parallelized SSIM algorithm, on a 2.8GHz Dual Core PC with 4GB memory. Our algorithm takes 2 to 8 seconds to opti-

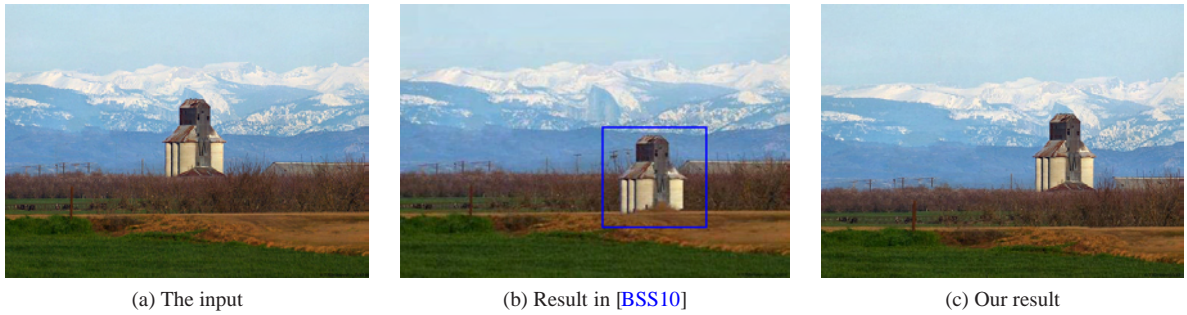


Figure 6: Comparison with the method in [BSS10]. (a) The input image in [BSS10]. (b) The result produced by the method of [BSS10]. Note that, the building is originally located on the grassland in the distance. However, its location is changed to the yellow land near the viewpoint in the result of [BSS10]. This in fact changes image semantics. (c) Our result.

mize the composition of photos of different sizes, if only a pair of seams is processed each time. However if we carve out and insert several seams each time, it takes around 1 sec. The algorithm can be further accelerated by transplanting it onto GPUs and exploiting the parallel computing power of modern Graphics card.

Limitation. We use seam carving as the basic operation for improving image compositions. Seam carving will break dense structures when the input image has complex structures and the seams will pass through them inevitably. It is one drawback of our algorithm. We show a failure result for an input image shot in the Grand Canyon, as shown in Figure 8. Since the fence runs through the image from left to right, the barbed wire is broken if too many vertical seams passing through it are removed, even though edge similarity is considered in our edge-based SSIM measure. To handle this problem, it will be helpful to exploit other structure-preservation image retargeting methods [GLS*09, ZCHM09].

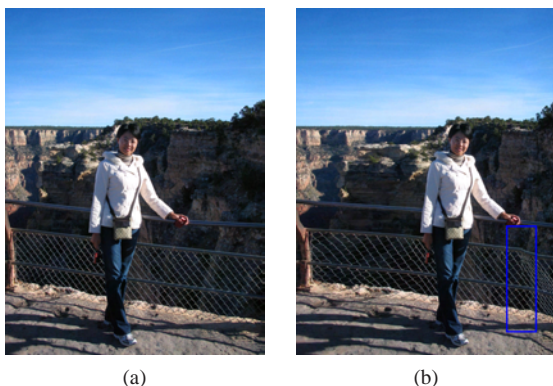


Figure 8: A failure case. The barbed wire is broken in the result (b). See the blue rectangle.

6. Conclusions

We have presented a new algorithm for improving image compositions by optimizing a unified objective model of composition aesthetics and image similarity. An edge-based measure of structural similarity that compares the optimized image and the original one is used. With the similarity constraint, our algorithm ensures visual similarity, and to some extent, semantic consistency between the optimized images and the results. By searching the maximum of the objective model, we are able to generate the composition improved images with nearly unperceivable visual distortions. Our algorithm is simple, intuitive, and easy to implement.

Since our algorithm mainly concentrates on how to ensure image similarity in the process of improving composition, we now compute composition aesthetics only under the guidance of rule of thirds. Professional photographers, however, may have various disciplines and usually take photos according to their rich experiences. In order to achieve this, previous techniques have used a learning model to map the aesthetic features to user input image attractiveness or adopted more photography guidelines for compensating rule of thirds. We intend to integrate their methods of evaluating composition aesthetics into our framework in future. In addition, it would be interesting to combine composition optimization with those tone adjustment and color harmonization techniques for making photographs by non-professional photographers professional and attractive.

7. Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable and constructive comments. This work was supported in part by the National Science Foundation of China under Grants 61073098 and 61021062, and the National Fundamental Research Program of China (2010CB327903).

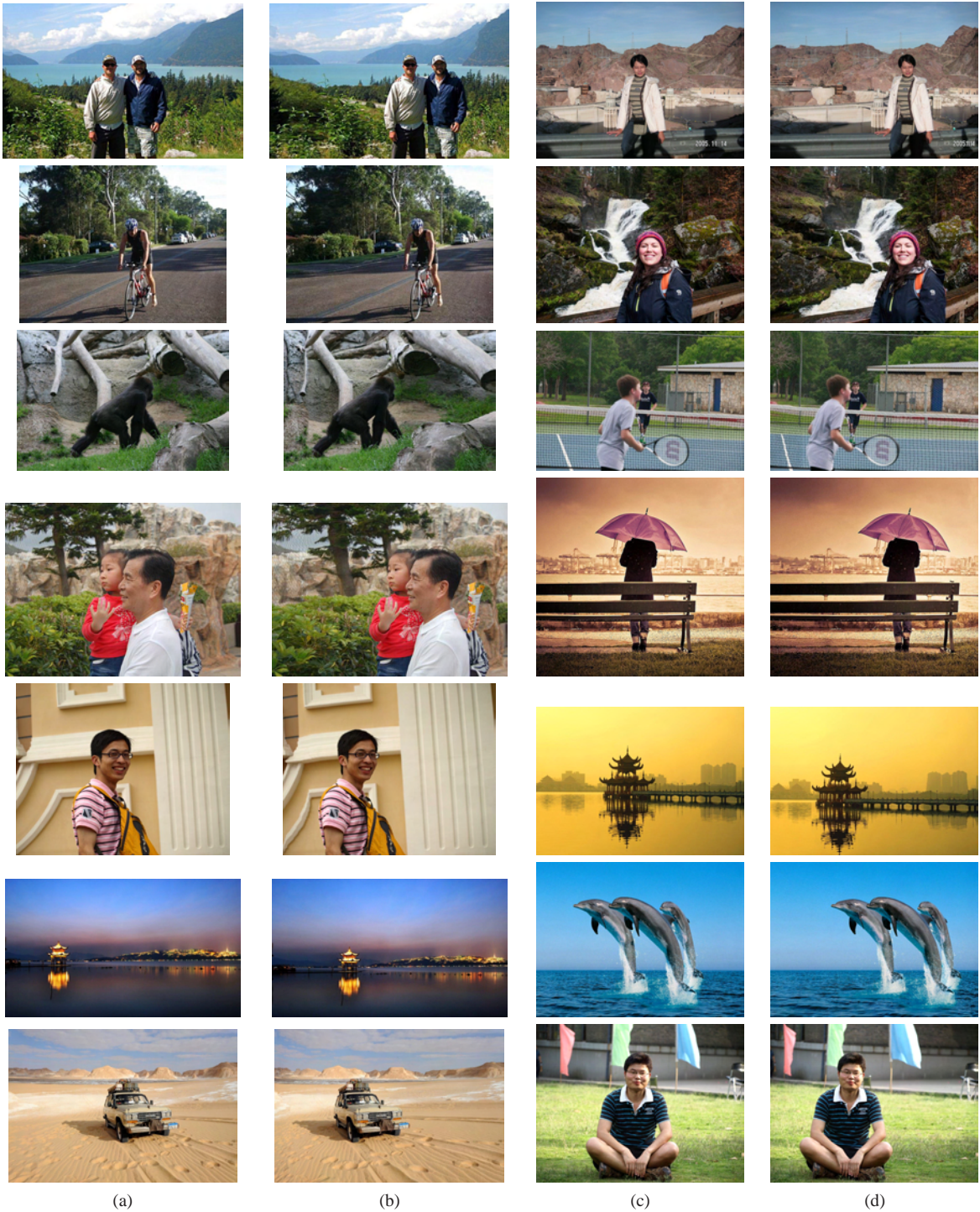


Figure 7: More results. (a), (c) The input images. (b), (d) Our results.

References

[AS07] AVIDAN S., SHAMIR A.: Seam carving for content-aware image resizing. *ACM Transactions on Graphics (Siggraph)* 26, 3

(2007), 267–276.

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A.,

- GOLDMAN D. B.: Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Siggraph)* 28, 3 (2009), 1–11.
- [BSS10] BHATTACHARYA S., SUKTHANKAR R., SHAH M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In *ACM Multimedia* (2010), pp. 271–280.
- [CCT*09] CHEN T., CHENG M.-M., TAN P., SHAMIR A., HU S.-M.: Sketch2photo: Internet image montage. *ACM Transactions on Graphics (Siggraph Asia)* 28, 5 (2009), 124:1–10.
- [COSG*06] COHEN-OR D., SORKINE O., GAL R., LEYVAND T., XU Y.-Q.: Color harmonization. *ACM Transactions on Graphics (Siggraph)* 25, 3 (2006), 624–630.
- [CZM*11] CHENG M.-M., ZHANG G.-X., MITRA N. J., HUANG X., HU S.-M.: Global contrast based salient region detection. In *IEEE CVPR* (2011), pp. 409–416.
- [GLS*09] GUO Y., LIU F., SHI J., ZHOU Z.-H., GLEICHER M.: Image retargeting using mesh parametrization. *IEEE Transactions on Multimedia* 11, 4 (2009), 856–867.
- [HZZ11] HUANG H., ZHANG L., ZHANG H.-C.: Arcimboldo-like collage using internet images. *ACM Transactions on Graphics (Siggraph Asia)* 30, 6 (2011), 155:1–7.
- [IK01] ITTI L., KOCH C.: Computational modelling of visual attention. *Nature Reviews Neuroscience* 2, 3 (2001), 194–203.
- [KTJ06] KE Y., TANG X., JING F.: The design of high-level features for photo quality assessment. In *IEEE CVPR* (2006), pp. 419–426.
- [LCODL08] LEYVAND T., COHEN-OR D., DROR G., LISCHINSKI D.: Data-driven enhancement of facial attractiveness. *ACM Transactions on Graphics (Siggraph)* 27, 3 (2008).
- [LCWCO10] LIU L., CHEN R., WOLF L., COHEN-OR D.: Optimizing photo composition. *Computer Graphics Forum (Eurographics)* 29, 2 (2010), 469–478.
- [LG05] LIU F., GLEICHER M.: Automatic image retargeting with fisheye-view warping. In *ACM UIST* (2005), pp. 153–162.
- [LJW10] LIU L., JIN Y., WU Q.: Realtime aesthetic image retargeting. In *Eurographics Workshop on Computational Aesthetic in Graphics, Visualization, and Imaging* (2010), pp. 1–8.
- [LWT11] LUO W., WANG X., TANG X.: Content-based photo quality assessment. In *IEEE ICCV* (2011), pp. 2206–2213.
- [NOSS09] NISHIYAMA M., OKABE T., SATO Y., SATO I.: Sensation-based photo cropping. In *ACM Multimedia* (2009), pp. 669–672.
- [Pet04] PETERSON B. F.: *Learning to see creatively*. Amphoto Press, 2004.
- [RGWZ08] REN T., GUO Y., WU G., ZHANG F.: Constrained sampling for image retargeting. In *IEEE ICME* (2008), pp. 1397–1400.
- [RSA09] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Multi-operator media retargeting. *ACM Transactions on Graphics (Siggraph)* 28, 3 (2009).
- [SAD*06] SANTELLA A., AGRAWALA M., DECARLO D., SALESIN D., COHEN M.: Automatic thumbnail cropping and its effectiveness. In *ACM CHI* (2006), pp. 771–780.
- [SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *IEEE CVPR* (2008), pp. 1–8.
- [SLBJ03] SUH B., LING H., BEDERSON B. B., JACOBS D. W.: Automatic thumbnail cropping and its effectiveness. In *ACM Conf. User Interface and Software Technology* (2003), pp. 95–104.
- [WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.
- [WTSL08] WANG Y. S., TAI C.-L., SORKINE O., LEE T.-Y.: Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics (Siggraph Asia)* 27, 5 (2008), 118:1–118:8.
- [ZCHM09] ZHANG G.-X., CHENG M.-M., HU S.-M., MARTIN R. R.: A shape-preserving approach to image resizing. *Computer Graphics Forum (Pacific Graphics)* 28, 7 (2009), 1897–1906.
- [ZFL*10] ZHOU S., FU H., LIU L., COHEN-OR D., HAN X.: Parametric reshaping of human bodies in images. *ACM Transactions on Graphics (Siggraph)* 29, 4 (2010), 1–10.