# Aesthetic Image Enhancement by Dependence-Aware Object Re-Composition

Fang-Lue Zhang, Miao Wang, Shi-Min Hu, *Member, IEEE*

***Abstract*—This paper proposes an image enhancement method to optimize photo composition, by rearranging foreground objects in the photo. To adjust objects' positions while keeping the original scene content, we first perform a novel structure dependence analysis on the image to obtain the dependencies between all background regions. To determine the optimal positions for foreground objects, we formulate an optimization problem based on widely used heuristics for aesthetically pleasing pictures. Semantic relations between foreground objects are also taken into account during optimization. The final output is produced by moving foreground objects, together with their dependent regions, to optimal positions. The results show that our approach can effectively optimize photos with single or multiple foreground objects without compromising the original photo content.**

***Index Terms*—image enhancement, photo composition, region dependence**

## I. Introduction

In recent years, the rapid development of digital photography has fostered demand for image enhancement techniques. Much work has been devoted to converting problems of visual quality enhancement into computational ones. Such methods can greatly improve photo quality based on global visual features such as tone [10] and clarity [30][23]. In the theory of visual psychology, human aesthetic judgments are mainly dependent on object-related cognition and processing [26], and the geometric structure of the entire image is also an important aesthetic element [3]. However, the above methods do not support object-level manipulation to improve aesthetic structural qualities.

In aesthetic evaluation of images, *composition* considers object relationships and geometric structure, which is one of the most influential aesthetic factors [22][46][18]. Recently, a few researchers have attempted to use photographic composition rules in image processing algorithms to improve the aesthetic quality. [31] first formulated the composition improvement problem as an optimization framework, and used cropping-and-retargeting operations to achieve high quality composition results. However, this method can potentially lose background information and may fail when objects are too large or too close to the border. Directly repositioning objects can avoid this problem. In [7], objects can be moved to new positions suggested by a learning-based algorithm. Retargeting approaches were used in [32][19] to optimize objects' positions. However, the position of each object is determined separately, and there is no consideration of the global layout of the objects as a group. In this work, we focus on how to move foreground objects to positions to produce a result with greater aesthetic quality. There are two main challenges: (a) how to move the foreground objects without causing inconsistency with the background; (b) how to express as a computational problem the desire to find optimal positions of salient objects taken as a group, taking into account inter-object relationships. We need an effective method to analyze the scene structure and decide which regions should be moved together with the foreground objects, and also a computational model to determine the best layout, which considers both semantic and geometric relations between different objects.

We propose a system to rearrange foreground objects and optimize photo composition. Our system has two main components: region dependence analysis and object position optimization. In dependence analysis, we use graph cuts based methods to optimize the dependence between over-segmented regions and the extracted objects, allowing us to determine which regions should be moved with each object. Using this approach, the scene structure around objects can be retained during repositioning. To determine the new object positions, we solve an optimization problem based on a set of well-

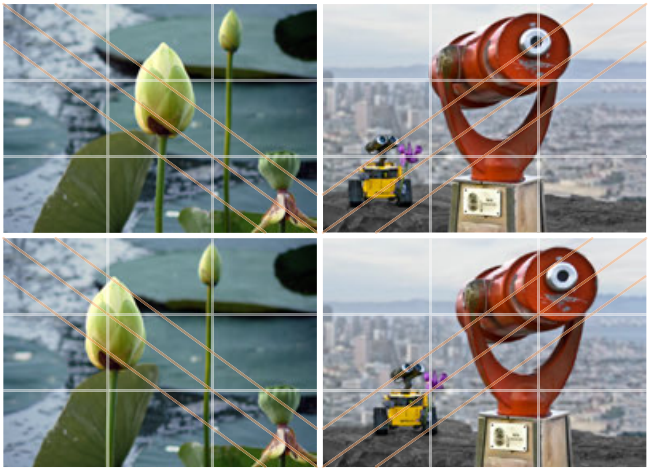


Fig. 1. Photo composition optimization. The images in the lower row are optimized results, where the photographic composition rules are satisfied: the white guide lines are based on the 'rule of thirds'; the orange guide lines are based on the 'diagonal frame'.



Fang-Lue Zhang, Miao Wang, and Shi-Min Hu are with the Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: z.fanglue@gmail.com; wangmiaoxdu@gmail.com; shimin@tsinghua.edu.cn).

Shi-Min Hu is the corresponding author.

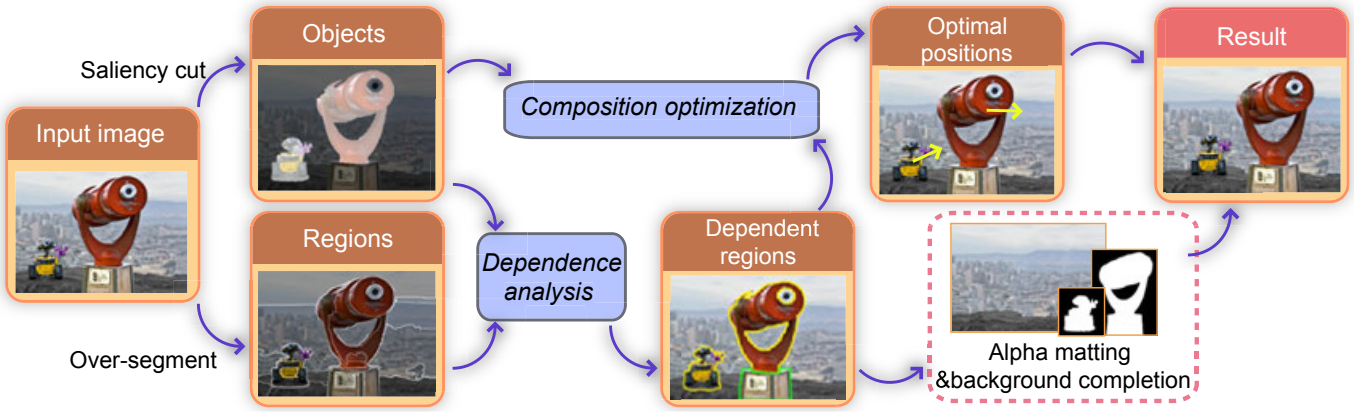



Fig. 2. Flowchart. We first extract objects and segment the image into regions, then perform a dependence analysis on them. Objects' optimal positions are calculated by composition optimization. Objects together with their dependent regions are composited at optimal positions on the completed background.

known photographic composition rules. The final result is produced by placing each object with its dependent regions at optimal positions on the completed background.

The main contributions of our work are:

- A novel method to analyze the dependence between regions and objects in images, which considers both photographic and psychological impact.
- A formulation as an optimization problem for global object layout improvement, taking into account inter-object relations.

## II. Related Work

Aesthetic assessment of photos has been investigated in several previous works, based on global and local visual features [42][24][29], and quantifiable aesthetic principles [14][28]. Yet, these methods only provide an overall evaluation of photos, instead of guidance for geometry structure adjustment to improve the aesthetic quality, which is the focus of this work.

To improve the visual aesthetics of digital photographs, some researchers have considered how to manipulate image contents following aesthetic and psychological principles. Santella et al. [41] and Nishiyama et al. [35] performed cropping on the original photos to find a best output based on users' attention. [31] proposed a composition optimization approach using cropping and retargeting operators. However, cropping based methods can potentially lose background information and can fail for large objects. Another attempt to improve photo composition by manipulating foreground objects is found in [7], where several photographic composition rules were used as guidance for placing objects at their best positions. However, the relations between different objects are not taken into account , and there is no guarantee of keeping semantic information presented in the scene. Our work *does* consider such issues to get a global optimal layout.

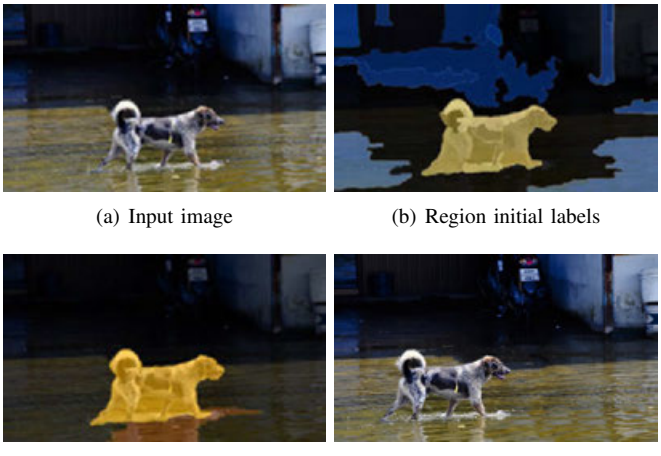
Other approaches to improve aesthetic quality of photos also exist. In-camera systems that automatically adjust the camera settings to satisfy compositional rules have been developed [1], and some aesthetic features such as depth-of-field can be automatically controlled by the in-camera system [4]. However, these in-camera systems cannot improve the photo composition after shooting. Recently, Merrell et al. [34] presented a furniture layout guiding system based on some aesthetic rules, but it cannot deal with the photo composition problem we focus on.

Image enhancement and editing are key tools in computer graphics. Early work of Porter and Duff [38] used an alpha matte to composite objects. Recent advances in alpha matting [44] have made it possible to generate more natural and visually pleasing results. Poisson blending [37] and its variations [49][8][48] reduce color mismatching by using gradient domain computations. Farbman et al. [16] showed how to achieve similar composition results efficiently. Various pixel- and patch-based approaches [47] also exist which underpin many applications like image reshuffling and inpainting [13][5], [39][43]. Shape aware image editing methods enable object-level operations [11][20][50]. These works provide powerful interactive tools to manipulate image content, but they do not consider aesthetics, which have the potential to guide amateur users in achieving better visual results.

Related research is also found in the field of computer vision. Unsupervised image segmentation approaches like [15][51][21] provide the foundation for image structure analysis. Saliency detection methods [12], have been integrated into image segmentation methods [40] which extract foreground objects with high visual attention automatically. However, relationships between segmented regions are not extracted in these methods. If we perform operations only on certain regions, a main problem is that the underlying semantic structure of the scene may be damaged as shown in Figure 4(d) and (h).

## III. Overview

Given an input image, we adjust its composition by moving objects to produce a better layout, while keeping the original frame and background. The algorithm framework is



(a) Input image (b) Region initial labels

(c) Object and dependent regions (d) Composition optimized result

Fig. 3. Region dependence analysis. Given an input image (a), analysis generates initial labels for regions belonging to objects and background (b). After multi-label graph cuts optimization, regions dependent on each object are determined (c); the yellow region is the saliency-cut result representing the dog and the orange area shows its dependent regions. The layout optimization result is shown in (d). The water splash and reflection in the water have moved with the dog, retaining the original semantic local structure.

shown in Figure 2. First, we extract clear foreground objects using a saliency cut [12] method. Then over-segmentation is performed to divide the image into regions. We analyze the dependence relationships between foreground objects and background regions using a novel method based on multi-label graph cuts, and determine those dependent regions that should be moved together with associated objects during repositioning. Section IV explains the structure dependence analysis in detail.

Knowing the dependence between foreground objects and background regions allows the algorithm to retain the semantic structure when designing the new composition. To determine the best layout of foreground objects, we formulate the aesthetic layout for multiple objects into an optimization problem. The optimization considers not only aesthetic rules, but also inter-object relations and connections between foreground objects and the background. In Section V, we show how we formulate the layout problem in terms of optimization. Finally we use alpha matting [27] to obtain a precise region with opacity value for each object and its dependent regions, and place them at the optimal positions in the background completed by PatchMatch [5].

## IV. Photo Structure Analysis

Changing the position of an object can damage the structure of the original scene. In existing image reshuffling work [5], semantic information relating background components and target objects was not considered. In this section, we describe how dependence analysis is performed to determine dependent regions which should move together with objects.

### A. Preprocessing

First, to understand the structural contents of an image, and the relationships between them, we extract the foreground objects and make a fine segmentation to obtain regions as the input of our analysis algorithm. This is achieved through several automatic operations on the input photo.

***Foreground objects extraction*** Saliency detection methods can be combined with interactive segmentation methods to detect foreground objects from an image. We use the saliency-cut method proposed in [12] to extract major objects. In saliency-cut, pixels with high (or low) saliency values are labeled as foreground (or background), which are then passed to GrabCut [40]. The segmented foreground result is regarded as a foreground object. We use saliency-cut to sequentially extract each foreground object using an iterative process. At each iteration, pixels belonging to the object extracted in the last iteration are set to background, and the saliency threshold for the pixels which will be set as foreground is reduced by a constant amount of 0.04, where saliency values range from 0 to 1.

***Pre-segmentation*** We use the automatic image segmentation method in [17] to divide the input image into over-segmented regions; these are taken as the basic structural elements of the image. The saliency value $s_k$ for each region $r_k$ is obtained by the *region contrast* method [12]. For the $i$th object, we group all regions that have more than half of their area covered by the object as the object regions. Regions with a saliency value smaller than some threshold $t$ are regarded as pure background.

### B. Features of regions for dependence analysis

To measure the degree of visual dependence between regions, each region needs to be quantified with proper visual features. Based on photography and psychology [46], the following features were carefully selected after multiple experiments.

*1) Acutance:* Foreground objects and regions closely related to them, i.e. physically close or semantically relevant to the objects, more readily draw people's attention because they have higher local contrast than the main background, or higher *acutance*. Acutance describes how quickly image information transfers at an edge, and high acutance results in sharp transitions and details with clearly defined borders. Based on a weighted average of second-order derivatives of pixels, we may measure the acutance of a region as

$$E_a = \mathcal{G}\left[\frac{1}{n}\sum_{i=1}^{n}\delta(i)D(i)\right] \quad (1)$$

in which $\delta(i) = 1$ if the second-order derivative $D(i)$ is larger than a threshold $t'$. We take $t' = 0.1$. $n$ denotes the total number of pixels in the region, $\mathcal{G}[*]$ is the Gaussian normalization function in [2].

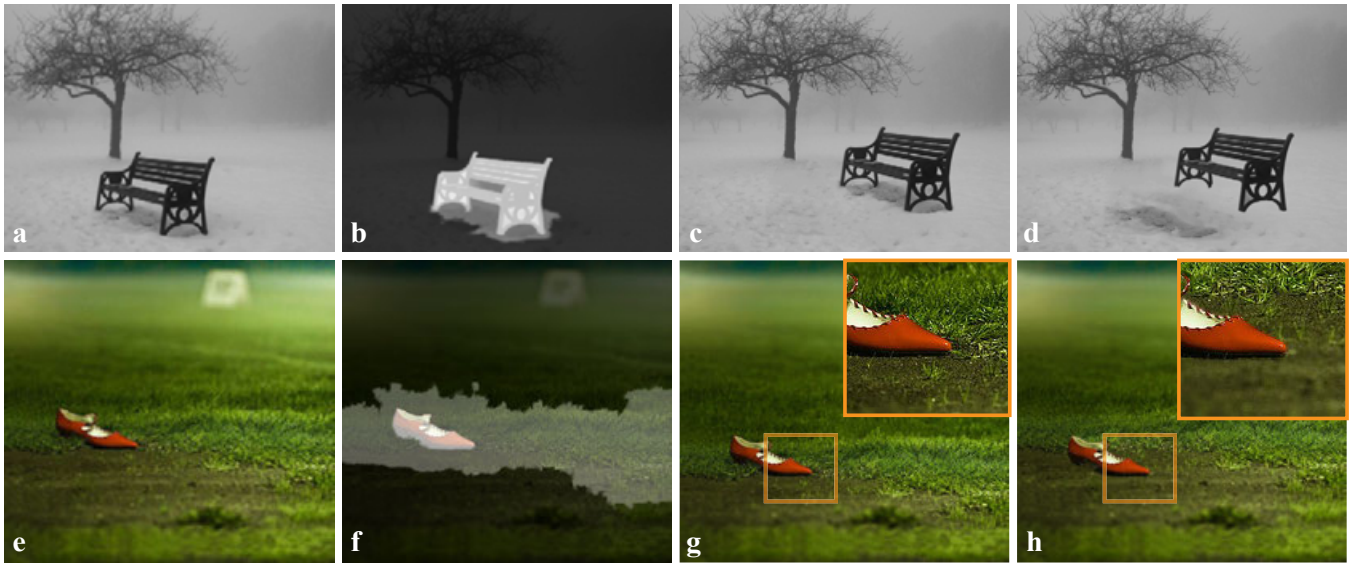



Fig. 4. Dependence analysis. Input images: (a), (e). In (b), (f), objects are shown as a pale mask, and dependent regions are shown as a gray mask. (c), (g) show results of moving the objects with their dependent regions. Moving objects alone may destroy the structure of the image: see (d), (h). In (d), the snow region dependent on the chair does not move with it, which fails to keep semantic information consistent. In (h), a clear shoe is put on a blurred ground, damaging the overall depth-of-field structure.

*2) Sharpness:* Photos may have a greater or lesser depth-of-field (DOF); regions in focus have higher sharpness of details. If a focused object does not move with those surrounding regions which share the same depth, this will cause damage to the DOF structure of the photo, as shown in Figure 4(h). Thus, sharpness is an important feature. Regions with greater sharpness typically have more energy in the high frequency range of the Fourier spectrum of the image. Therefore, inspired by [33], we use the the ratio between higher and lower frequency-band energy to measure sharpness,

$$E_s = \mathcal{G}\left[\sum_{(u,v)\in\mathcal{F}_H} F(u,v) / \sum_{(u,v)\in\mathcal{F}_L} F(u,v)\right] \tag{2}$$

$$\mathcal{F}_H = \{(u,v) \mid \beta W < |u-u_0| \le \alpha W, \beta H < |v-v_0| \le \alpha H\}$$

$$\mathcal{F}_L = \{(u,v) \mid |u-u_0| \le \beta W, |v-v_0| \le \beta H\}$$

where $W$ and $H$ are the width and height of the image, $\mathcal{F}_H$ is the high-frequency band, $\mathcal{F}_L$ is the low-frequency band, and $(u_0, v_0)$ is the central frequency. In our experiments, $\alpha = 0.4, \beta = 0.2$.

*3) Harmony between main colors:* When moving the objects, the region surrounding them should be harmonious with them, making the objects more consistent and coordinated with surrounding elements. Thus, we add harmony between main colors of adjacent regions to help decide which regions should move along with objects. We use the color-harmony model proposed by Ou et al. [36] for color combinations. Given two colors in CIELAB space, the *harmony* may be calculated as

$$CH = H_C + H_L + H_H \tag{3}$$

where

$$H_C = 0.04 + 0.53 tanh(0.8 - 0.045 \triangle C)$$

$$H_L = H_{Lsum} + H_{\triangle L}$$

$$H_H = E_{C1} * (H_{S1} + E_{Y1}) + E_{C2} * (H_{S2} + E_{Y2})$$

For more details, see [36]. The range of the above score is from -5 to 5, so we define a color harmony distance between the two region's main colors as:

$$\triangle CH_{1,2} = exp\{(CH_{1,2} + 5)^2/2\} \tag{4}$$

### C. Dependence analysis by multi-label graph cuts

Objects in an image are not always independent on the background. As shown in Figure 4, if the connections between objects and certain background regions are broken, the semantic structure can be destroyed. We need to determine which regions have stronger ties with foreground objects, and which ones are more like background components. When moving an object, the regions strongly tied to the object should be moved with it, keeping their relative positions. Thus, all regions are classified as either a *dependent region* of a certain object or background.

To determine dependence relations, the following issues need to be considered. First, if there is a clear DOF layering structure, the regions sharing the same sharpness as the object should be set as dependent on it, as shown in Figure 4(e). Secondly, regions semantically related to objects, with higher acutance than other background regions, or having harmonious colors with the objects are also set as dependent. An example of this is the rough snow under the chair in Figure 4(a). To obtain the two kinds of *dependence*, we use the features in Section IV-B to describe each region, and measure the dependence in the feature space.

If a pair of regions $(r_i, r_j)$ has adjacent pixels, we mark them as neighbors. The set of all such neighbor pairs is denoted as $\mathcal{N}$. Thus, considering the labels (denoted by $\mathcal{L}$) of reference regions belonging to objects and the low-saliency background as possible labels for the other regions, labels are assigned by cost optimization. Each target region $r \in R$ is given a label from $\mathcal{L}$ according to feature distances between the region and its label, and the desire for neighbors to have a common labeling. The energy function is defined as:

$$E(L) = \sum_{r_i \in \Omega} D_{r_i}(L_i) + \sum_{(r_i, r_j) \in \mathcal{N}} T_{(r_i, r_j)}(L_i, L_j) \quad (5)$$

in which $\Omega$ is the set of all the regions, $L_i \in \mathcal{L}$. The label energy term is defined by the distance in feature space:

$$D_r(L) = \{[E_s(r) - E_s(L)]^2 + [E_a(r) - E_a(L)]^2 + \triangle CH_{r,L}\}^{\frac{1}{2}} \quad (6)$$

Neighbor regions with similar features should be more likely to have the same label, so:

$$T_{(r_i, r_j)}(L_i, L_j) = \begin{cases} 0, & \text{if } L_i = L_j \\ D_{r_i}(r_j), & \text{otherwise} \end{cases} \quad (7)$$

A weighted undirected graph $G = (V, E)$ is constructed over all regions (see Figure 5). The nodes $V$ correspond to the regions, and for neighbors $(i, j) \in \mathcal{N}$, we add edge $e_{ij}$ to $E$. The weight of $e_{ij}$ is $D_i(j)$. We set the fixed labels for regions in $\mathcal{L}$ as themselves. Because the distance measurements are not metric, we use $\alpha - \beta$ swap to optimize the multi-label graph cuts problem as proposed in [9]. Given $|\mathcal{L}|$ labels, it takes $|\mathcal{L}|^2$ iterations to perform the swap algorithm. In each iteration, a max-flow algorithm with a complexity of $O(|E||V|^2)$ in the worst case is performed, where $|E|$ is the number of edges, and $|V|$ is the number of regions. Thus, the overall complexity of the dependency analysis is $O(|E||V|^2|\mathcal{L}|^2)$ in the worst case. As Figure 3 shows, graph cuts optimization gives the semantically dependent regions for the foreground objects. The structure of the original photo can be preserved after moving the object with dependent regions.

## V. Object Layout Optimization

Good composition is obtained by carefully placing objects [22], [46], [18]. In this section, we convert the commonly used composition rules into computable measurements, and transform the layout problem into an optimization problem to find the optimal positions for objects.

### A. Layout optimization objective

To formulate our objective for layout optimization, aesthetic criteria for composition are used in combination with constraints and relations between objects. In [31], the *rule of thirds* and *visual balance* are used as guidelines for salient regions' positions to perform a best cropping-retargeting operation. We also adopt these two well known composition rules in our object layout evaluation. The diagonal rule [18] is another an important guideline for laying out multiple objects, and we also include it in our optimization model. In addition, as freely placing objects easily damages the global structure of the original image, various constraints are included to limit the changes and correlation between objects is used to maintain possible semantic relationships.

We normalize the positions of the objects in $[0, 1]^2$. The centroid of the most salient region of object $i$ is used as the object center $c_i$. The mass of the $i$th object is the number of pixels it contains, normalized by the total number of pixels in all objects, and denoted as $m_i$. Our object layout optimization objective for photo composition is built from the following terms:

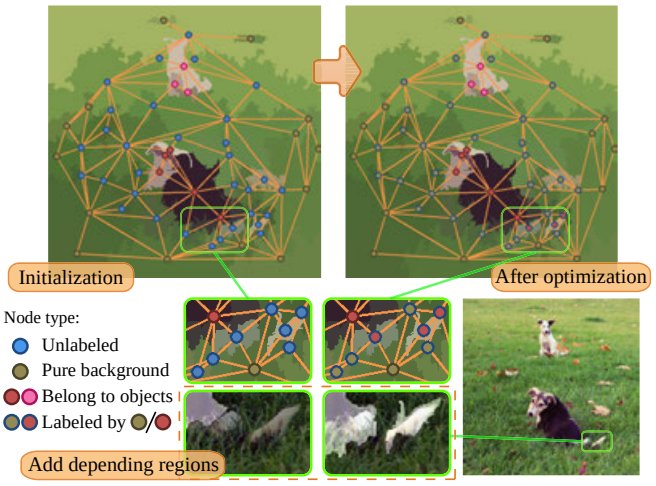



Fig. 5. Region graph. Left: initial graph. Right: result after graph cuts optimization. Below: masks of regions to move before and after adding the dependent regions.

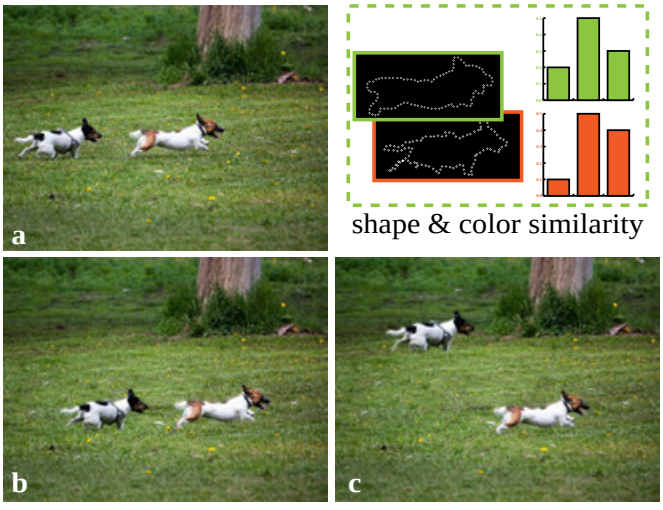



Fig. 7. Relevance between objects. (a) Original image. (b) Layout optimization result with object relevance constraint. (c) Layout optimization result without object relevance constraint. The semantic relation is lost.


(a) Distance from power points (b) Distance from diagonal lines (c) Visual balance (d) Relevance of the objects

Fig. 6. Composition rules and constraints for optimization. Orange lines: image diagonals. White lines: guides for the rules of thirds. Yellow circles: object centroids, size of circle determined by its mass. Dotted green lines: connection lines between objects.

*1) Distance from power points:* The *power points* [31] are the four intersections of the horizonal/vertical lines in the *rule of thirds*. To make the photo more appealing, the photographer is often advised to place key foreground objects on one of the power points. In normalized coordinates, the power points are $P = \{(\frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, \frac{1}{3}), (\frac{2}{3}, \frac{2}{3})\}$. The energy term for the power point distance is defined as:

$$D_P = \sum_i^n m_i \|c_i - P_j\| \tag{8}$$

where $P_j$ is the nearest power point to the current position.

*2) Distance from diagonal lines:* According to the diagonal rule, important elements of the picture should be placed along these diagonals, and the line through two foreground objects should be also along one of the diagonals [18]. The associated energy term is calculated by the sum of all the distances between line segments $\ell_{ij}$ connecting each pair of objects and the two diagonal lines $L_1$ and $L_2$.

$$D_L = \sum_i^n \sum_{j \neq i}^n \frac{1}{2}(f_a(\ell_{ij}, L_1, L_2) + f_d(\ell_{ij}, L_1, L_2)) \tag{9}$$

where

$$f_a(\ell_{ij}, L_1, L_2) = \frac{|\theta_1||\theta_2|}{4\pi^2}$$

$$f_d(\ell_{ij}, L_1, L_2) = \sqrt{2}d(\mathcal{M}_{ij}, L_1) \cdot \sqrt{2}d(\mathcal{M}_{ij}, L_2)$$

In the above equations, $\mathcal{M}_{ij}$ is the mid-point of $\ell_{ij}$, and $\theta_k$ is the angle between $\ell_{ij}$ and $L_k$. $f_d(\ell_{ij}, L_1, L_2)$ measures the distance between $\ell_{ij}$ and the diagonals $L_1$ and $L_2$, and $f_a(\ell_{ij}, L_1, L_2)$ measures the angular distance between $\ell_{ij}$ and the diagonals. The term reaches a minimum value when the line segment is similar and close to one of the diagonals, and the maximum value when $\ell_{ij}$ is a vertical or horizonal line with equal distance from the two diagonals. Normalisation constants ensure values in $[0, 1]$.

*3) Visual balance:* *Visual balance* is a well known aesthetic criterion in art. The method proposed in [31] is adopted here, and we use the distance from center of mass of all objects to the image center $C$ $(0.5, 0.5)$ as the visual balance value. Let $\mathcal{M}_m = \sum_i^n m_i c_i$ denote the center of mass. This term is defined as:

$$D_V = exp\{-\frac{1}{2\sigma} d^2(C, \mathcal{M}_m)\} \tag{10}$$

where $\sigma = 0.2$.

*4) Relevance of objects:* Changing objects' relative positions may damage semantic information in the image (see Figure 7). To maintain the semantics, relative positions of relevant objects should be kept consistent when moving them. Objects with similar shapes or in the same category are often used as highly relevant foregrounds when people take photos [18]. We use shape similarity and color distribution similarity to measure relevance. The shape similarity $S_{shape}(i, j)$ is measured by shape context [6]. In terms of color similarity measurement, we first quantize each color channel into 12 values in $L * a * b$ color space giving $K = 12^3$ colors, then calculate objects' histograms $H$ in the $L * a * b$ color space. Next we compare each pair of objects' histograms $H_i, H_j$ using $\chi^2$-distance to obtain the score $S_{color}$:

$$S_{color}(i, j) = \frac{1}{2} \sum_{k=1}^{K} \frac{[H_i(k) - H_j(k)]^2}{H_i(k) + H_j(k)} \tag{11}$$

The total similarity score $S$ is calculated as:

$$S(i, j) = \lambda * S_{shape}(i, j) + (1 - \lambda) * S_{color}(i, j) \tag{12}$$

where $\lambda$ is a tuning parameter, set to 0.5 in our experiments. Let $\triangle_{i,j}$ denote the relative position between original positions of $i$ and $j$, and let $\triangle'_{i,j}$ be the changed relative position. The energy term with respect to change of relative positions is:

$$R(i, j) = S(i, j) \|\triangle_{i,j} - \triangle'_{i,j}\| \tag{13}$$

*5) Constraints and penalty:* Given the above four energy terms, free repositioning can still lead to results compromising the original scene structure. Thus, we add a penalty term to ensure the final positions are the nearest optimal solutions to the initial layout. For the $i$th object, the penalty value is:

$$\mathcal{P}'_i = \frac{1}{\alpha} \|c'_i - c_i\|, \quad \mathcal{P}_i = \underline{\mathcal{P}'_i}$$

Where $\underline{\mathcal{P}'_i}$ means taking integers of $\mathcal{P}'_i$ downwardly. We set $\alpha = 1/3$. $\mathcal{P}_i$ adds 1 for each additional $K$ in $\mathcal{P}'_i$. Sometimes, an object and its dependent regions may reach the image boundary, in which case it cannot be moved freely as there is insufficient information to complete the object and the dependent regions. For example, see Figure 8(b) and (d).

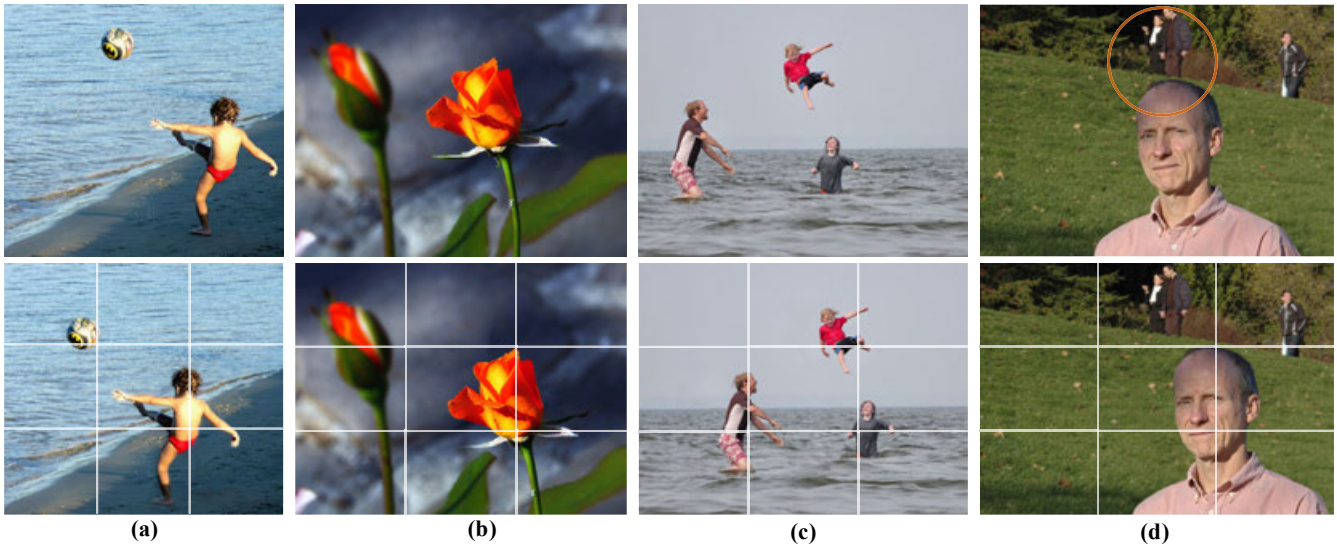



Fig. 8. Optimization results. Above: input images. Below: our optimized output images.

We have to limit the motion allowed in such situations, the foreground objects' vertical positions should not be higher than their original vertical positions.

### B. Layout optimization

Given the above energy terms, the optimization objective is:

$$E = D_P + D_L + D_V + \sum_{i}^{n} \mathcal{P}_i + \omega \sum_{i}^{n} \sum_{j \neq i}^{n} R(i,j) \quad (14)$$

The parameters for the objective function are the $x$ and $y$ coordinates of the centers of interest for all objects, so there are $2n$ variables each of which must lie in [0,1] in the normalized image coordinate system. The weight $\omega$ controls the impact of the relation between objects. A larger $\omega$ makes the relative positions of the objects change less. The default value for $\omega$ is 1. The heuristic method *particle swarm optimization* [25] is adopted to search for the optimal solution. In PSO, the worst-case running time complexity is $O(nm)$, where $n$ is the number of particles, and $m$ is the maximum iteration times. We use $n = 1000$ and $m = 100$ in our experiments. As Figure 1 shows, our optimization method can make the objects' positions better agree with the composition rules, improving the aesthetic quality.

***Generating output*** We calculate the alpha value of each object with its dependent regions to obtain a precise region mask as well as its opacity, using the method in [27]. Then we use *content aware fill* method in Adobe Photoshop to complete the background. The final result is produced by linear combination of pixel values of each foreground object in its new position, with its dependent regions, and the background.

## VI. Discussion and User Study

A number of examples are presented in this section to demonstrate the performance of our approach. All examples were tested on a PC with a Core 2 Duo CPU at 2.66GHz and 4GB RAM. Dependence analysis takes about 2s and position optimization takes 0.2s–0.6s for an 800 $\times$ 640 image.

The optimal solution of Equation (14) balances all energy terms, avoiding mechanical results that can otherwise appear as a result of single target optimization. In Figure 8(a), under the influence of the power point distance, diagonal line distance and visual balance term, the football is placed slightly above the power point, instead of exactly on it, which gives a better visual balance. Figure 8(c) shows a result with multiple objects, where the splashes around the


Fig. 9. Comparison with the method in [31].

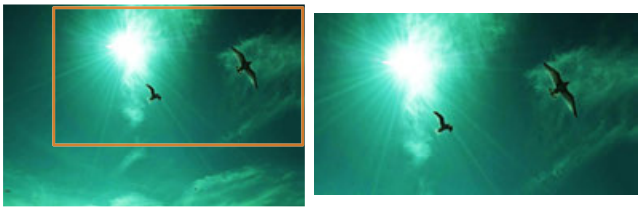


Fig. 10. Original image(Left) and optimized cropped frame(Right).

man and the boy are detected as dependent regions, and the global layout is more balanced after optimization. Sometimes, amateur photographers make composition mistakes like Figure 8(d), where two people in the background look to be standing on the head of the foreground man. Since the regions with the two people are not dependent on the foreground man because they have different sharpness and acutance, our method can deal with this kind of problem, improving the composition.

Our approach preserves the original scene structure to the degree possible. The semantic information in an image includes not only the relation of the background with objects, e.g. reflections on water, or regions sharing the same depth and focus, but it also includes the relations between objects. As shown in Figure 3 and Figure 4, moving the dependent regions together with an object produces more natural results. The effect of object relevance is shown in Figure 7. The two dogs have similar shapes, with a relatively large value for the relevance term in Equation (14). Thus, in the optimal solution, the chasing relation between these two dogs is retained. Additional results are shown in Figure 14 and supplemental materials.

Figure 9 compares our results with the crop-resizing photo composition method [31]. When objects are too close to the image border, it is difficult to improve compositions by cropping or resizing, but our method can automatically move such objects to a better position, improving the composition.

Our optimization framework also supports aesthetic cropping. Taking the top left corner of the cropping window and the width and height as the optimization space, we can find the optimal composition again using Equation (14). An example is given in Figure 10, where the objects have a more pleasant layout in the new frame.

***User Study*** A user study was performed to evaluate our method. Forty five pairs of photos were prepared, the original and our modified output; these were randomly placed next to each other. We invited twenty participants, 90% of them had no expertise in photography, and we did not tell them anything about the composition rules we used to optimize the photos. To eliminate bias, the participants were selected from different age and gender groups. There were 5 males and 5 females in both the group of age 18-30 and the group of age 31-45. They were asked to assign an integer rank from -3 to 3, to indicate how much more pleasing one's composition of the objects was than the other's. A positive score meant the photo on the right was more appealing than the one on the left and vice versa. The user study outcome is shown in Figure 11. It can be seen that our method improves the aesthetic quality effectively, with 91% percent of images judged to be improved to differing extents. More details for the user study are provided in the supplemental materials.

***Limitations*** We mainly focus on the layout of the foreground object positions. There are some other aspects of composition and scene structure in photography, such as the guiding lines/shapes, which are hard to detect and find a uniform formalized definition, and we do not integrate them into our optimization framework. This introduces limitations. As in Figure 12(top), there are abstract guiding lines formed by the tall-shaped front bird itself and the row of small birds, but the structure is destroyed by moving the large bird to the nearest power point. If the composition is not formed by the layout of objects, our method cannot improve the aesthetic quality either, e.g., a photo whose composition is formed by the lines/curves in the frame. It also introduces limitations when some semantic information of objects affects the composition, e.g., we cannot guarantee that the animal/person will face to the appropriate side after optimization.

There are some specific cases we may fail using the unified optimizing framework. As in Figure 12(bottom), moving the

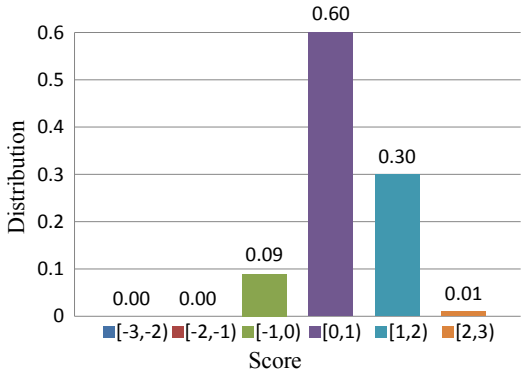



Fig. 11. The average score distribution of photos in user study.

(a) Input image (b) result (score: **-0.15**)

(c) Input image (d) result (score: **-0.85**)

Fig. 12. Examples received negative scores in the user study.

Fig. 13. Photos cannot be optimized by our approach. (Left) A photo with too complicated background and mutual occluded objects. (Right) An example for missing important part of foreground objects.

only foreground object which takes up a large portion of the image disrupts the visual balance. Sometimes, foreground objects have semantic relationship which cannot be detected by comparing visual features as in section V.A.4). How to improve their compositions is also beyond the capability of our method. In the future work, we need add more specific strategies, such as handling scenes with one large object, avoiding risks of collision between a lot of small objects, etc.

Our method is dependent on the object extracting results by the saliency-cut. Scattered background and occlusion between objects can lead to failures in objects extraction and dependence analysis. Requisite background completion and alpha matting may also fail to produce pleasing results in those cases. One such example is shown in Figure 13(left). Some amateur photographers make bad composition because the objects miss important parts, like the Figure 13(Right). We can not optimize them either, because there is not enough information to complete the objects.

## VII. Conclusion

We have proposed an automatic approach to optimize photo composition, based on repositioning foreground objects in the photo frame. The new approach includes two key components, structure dependence analysis and layout optimization, which enables the algorithm to keep the scene structure around objects being moved and to find the best position for each object. This approach improves the aesthetic quality of photos while preserving background information as well as the geometry of the original frame and each object. The user study shows that our automatic photo composition optimization method is effective in most cases. In the future work, further issues in photographic composition, such as simplicity, viewpoint, guiding lines, etc. will be explored; and more factors which influence aesthetic subjective sensation, like affections [45], will also be considered.

## Acknowledgments

The authors would like to thank the associate editor and all the reviewers for their helpful comments. This work was supported by the National Basic Research Project of China (Project Number 2011CB302205), the Natural Science Foundation of China (Project Number 61120106007), the National High Technology Research and Development Program of China (Project Number 2012AA011802) and National Significant Science and Technology Program (Project Number 2012ZX01039001-003).

## References

[1] R. Abdullah, M. Christie, G. Schofield, C. Lino, and P. Olivier, "Advanced composition in virtual camera control," in *Smart Graphics*. Springer, 2011, vol. 6815, pp. 13–24.

[2] S. Aksoy and R. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001.

[3] R. Arnheim, *Art and visual perception: a psychology of the creative eye*. University of California Press, 1969.

[4] S. Banerjee and B. Evans, "In-camera automation of photographic composition rules," *IEEE Trans. Image Processing*, vol. 16, no. 7, pp. 1807–1820, 2007.

[5] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: a randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, pp. 24:1–24:11, 2009.

[6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE TPAMI*, vol. 24, no. 4, pp. 509–522, 2002.

[7] S. Bhattacharya, R. Sukthankar, and M. Shah, "A holistic approach to aesthetic enhancement of photographs," *ACM Trans. on Multimedia Computing, Communications, and Applications*, vol. 7S, pp. 21:1–21:21, 2011.

[8] X. Bie, H. Huang, and W. Wang, "Free appearance-editing with improved poisson image cloning," *Journal of Computer Science and Technology*, vol. 26, no. 6, pp. 1011–1016, 2011.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222 –1239, 2001.

[10] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *IEEE CVPR*, 2011, pp. 97 –104.

[11] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Repfinder: Finding approximately repeated scene elements for image editing," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 83:1–8, 2010.

[12] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *IEEE CVPR*, 2011, pp. 409–416.

[13] M. J. Dahan, N. Chen, A. Shamir, and D. Cohen-Or, "Combining color and depth for enhanced image segmentation and retargeting," *The Visual Computer*, vol. 28, no. 12, pp. 1181–1193, 2012.

[14] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proc. ECCV*, 2006, pp. 7–13.

[15] A. Delong, A. Osokin, H. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," in *IEEE CVPR*, 2010, pp. 2173 –2180.

[16] Z. Farbman, G. Hoffer, Y. Lipman, D. Cohen-Or, and D. Lischinski, "Coordinates for instant image cloning," *ACM Trans. Graph.*, vol. 28, pp. 67:1–67:9, 2009.

[17] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, 2004.

[18] M. Freeman, *The Photographer's Eye: Composition and Design for Better Digital Photos*. The llex Press, lewes, England, 2007.

[19] Y. W. Guo, M. Liu, T. T. Gu, and W. P. Wang, "Improving photo composition elegantly: Considering image similarity during composition optimization," *Computer Graphics forum*, vol. 31, no. 7, pp. 2193–2202, 2012.

[20] H. Huang, L. Zhang, and H.-C. Zhang, "Arcimboldo-like collage using internet images," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 155:1–155:8, December 2011.

[21] T. Huang, Y. Tian, J. Li, and H. Yu, "Salient region detection and segmentation for general object recognition and image understanding," *Science China Information Sciences*, vol. 54, no. 12, pp. 2461–2470, 2011.

[22] P. Jonas, *Photographic composition simlified*. Amphoto Publishers, 1976.

[23] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski, "Image deblurring using inertial measurement sensors," *ACM Trans. Graph.*, vol. 29, pp. 30:1–30:9, 2010.

[24] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *IEEE CVPR*, vol. 1, 2006, pp. 419 – 426.

[25] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Conf. on Neural Networks*, 1995, pp. 1942 –1948.

[26] H. Leder, B. Belke, A. Oeberst, and D. Augustin, "A model of aesthetic appreciation and aesthetic judgments," *British Journal of Psychology*, vol. 95, no. 4, pp. 489–508, 2004.

[27] A. Levin, D. Lischenski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE TPAMI*, vol. 30, no. 2, pp. 228–242, 2008.

[28] C. Li, A. C. Loui, and T. Chen, "Towards aesthetics: a photo quality assessment and photo selection system," in *ACM MM'10*, 2010, pp. 827–830.

[29] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[30] Y. Ling, C. Yan, C. Liu, X. Wang, and H. Li, "Adaptive tone-preserved image detail enhancement," *The Visual Computer*, vol. 28, no. 6-8, pp. 733–742, 2012.

[31] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, no. 2, pp. 469–478, 2010.

[32] L. Liu, Y. Jin, and Q. Wu, "Realtime Aesthetic Image Retargeting," in *Proc. Eurographics Workshop on Computational Aesthetic in Graphics, Visualization, and Imaging*, 2010, pp. 1–8.

[33] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proc. ECCV*, 2008, pp. 386–399.

[34] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun, "Interactive furniture layout using interior design guidelines," *ACM Trans. Graph.*, vol. 30, pp. 87:1–87:10, 2011.

[35] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato, "Sensation-based photo cropping," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 669–672.

[36] L.-C. Ou and M. R. Luo, "A colour harmony model for two-colour combinations," *Color Research and Application*, vol. 31, pp. 191–204, 2006.

[37] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, pp. 313–318, 2003.

[38] T. Porter and T. Duff, "Compositing digital images," *ACM SIGGRAPH*, vol. 18, no. 3, pp. 253–259, 1984.

[39] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. ICCV*, 2009, pp. 151 –158.

[40] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[41] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," in *ACM CHI*, 2006, pp. 771–780.

[42] H. Sheikh, A. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2117 –2128, 2005.

[43] D. Wang, G. Li, W. Jia, and X. Luo, "Saliency-driven scaling optimization for image retargeting," *The Visual Computer*, vol. 27, no. 9, pp. 853–860, 2011.

[44] J. Wang and M. F. Cohen, "Image and video matting: a survey," *Found. Trends. Comput. Graph. Vis.*, vol. 3, pp. 97–175, 2007.

[45] X.-H. Wang, J. Jia, H.-Y. Liao, and L.-H. Cai, "Affective image colorization," *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1119–1128, 2012.

[46] E. A. Weber, *Vision, composition and photography*. Walter de Gruyte, 1980.

[47] C. Xiao, M. Liu, N. Yongwei, and Z. Dong, "Fast exact nearest patch match for patch-based image editing and processing," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 8, pp. 1122–1134, 2011.

[48] S. Xue, A. Agarwala, J. Dorsey, and H. Rushmeier, "Understanding and improving the realism of image composites," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 84:1–84:10, 2012.

[49] Y. Zhang and R. Tong, "Environment-sensitive cloning in images," *The Visual Computer*, vol. 27, no. 6–8, pp. 739–748, 2011.

[50] Y. Zheng, X. Chen, M.-M. Cheng, K. Zhou, S.-M. Hu, and N. J. Mitra, "Interactive images: Cuboid proxies for smart image manipulation," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 99:1–99:11, 2012.

[51] F. Zhong, X. Qin, and Q. Peng, "Robust image segmentation against complex color distribution," *The Visual Computer*, vol. 27, no. 6–8, pp. 707–716, 2011.

**Fang-Lue Zhang** received his BS degree from the Zhejiang University in 2009. He is currently a PhD candidate in Tsinghua University. His research interests include computer graphics, image processing and enhancement, image and video analysis and computer vision.

**Miao Wang** is currently a Ph.D. candidate at Tsinghua University, Beijing. He received his BS degree from Xidian University in 2011. His research interests include computer graphics, image processing and computer vision.

**Shi-Min Hu** received the PhD degree from Zhejiang University in 1996. He is currently a professor in the department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer-aided geometric design. He is associate Editor-in-Chief of The Visual Computer, associate Editor of Computer & Graphics and on the editorial board of Computer Aided Design. He is a member of the IEEE.

Fig. 14. Further results. In each group, the top image is the original, the middle image is the optimized result, and the bottom image indicates photography composition rules.