Optimizing Photo Composition

Ligang Liu[1][†]   Renjie Chen[1]   Lior Wolf[2]   Daniel Cohen-Or[2]

[1]Zhejiang University, China
[2]Tel-Aviv University, Israel

Figure 1: *Optimizing the aesthetics of the original photograph in (a) by our approach leads to the new image composition shown in (c). (b) shows the cropping result of the approach of [Santella et al. 2006]. The aesthetic scores are shown in (d). Our result in (c) obtains higher aesthetic score than (a). RT(rule of thirds), DA(diagonal), VB(visual balance), and SZ(region size) are components of the objective function.*

Abstract

Aesthetic images evoke an emotional response that transcends mere visual appreciation. In this work we develop a novel computational means for evaluating the composition aesthetics of a given image based on measuring several well-grounded composition guidelines. A compound operator of crop-and-retarget is employed to change the relative position of salient regions in the image and thus to modify the composition aesthetics of the image. We propose an optimization method for automatically producing a maximally-aesthetic version of the input image. We validate the performance of the method and show its effectiveness in a variety of experiments.

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

1. Introduction

Humans seek to achieve aesthetics in art. This goal is elusive since there is little consensus as to what makes one piece of art more aesthetic than another. Indeed, the judgment of aesthetics is subjective and involves sentiments and personal taste [MB98]. Despite the challenges, a new field called Computational Aesthetics has emerged. This area of research is concerned with the study of computational methods for predicting the emotional response to a piece of art, and in developing methods for eliciting and enhancing such impressions [Pet07, RPJJS07].

In this work, we focus on the aesthetics properties of image composition and employ rules that are well-known in the photography community. Such rules are routinely taught in professional courses and text-books [GS90, Kra05] as guidelines likely to increase the aesthetic appreciation of photographs.

Composition rules tell the photographer various aspects that he or she should consider when shooting a photograph. After the photograph is taken there is little that can be done to improve the composition of the picture, without laborious digital editing. Using commercial tools like Photoshop, one can crop the image, extract foreground objects and paste them back into the image. Photo touch-up is a routine for professional graphic designers, but not for the average amateur photographer.

Automating the process of aesthetic image adjustment requires the development of a computational aesthetic score

---

† Corresponding author: ligangliu@zju.edu.cn (Ligang Liu)

which represents the expected composition quality of a picture. We develop and formalize such a score based on a set of primary composition guidelines, including rule of thirds, diagonal dominance, visual balance, and size region. As far as we know, our work is the first attempt to incorporate the guidelines of diagonal dominance, visual balance, and size region in an automatic aesthetic score. As a result, tools for automatic photo touch-up may be defined as search problems.

In order to modify the composition of a given photograph, we employ a compound operator of crop-and-retarget. The cropping operator selects a subset of the image objects, then the retargeting operator adjusts their relative locations. The parameters of this dual operator are the coordinates of the crop window and the amount of inflation or deflation the image undergoes during the retargeting process. By searching for a combination of parameters that produces the image with the maximal aesthetic score, we generate an output image that is an improved version of the original one, and enable everyday photographers to create new photos with good composition from their own previously taken photos.

The specific contributions of our work include: i) identifying a set of composition rules, and implementing them computationally to allow a quantitative evaluation, ii) considering retargeting as an operator to change the relative position of salient regions in the image, and iii) facilitating an automatic image editing tool that enhances the aesthetics of a photograph, and everyday user's photography experience.

2. Background

Various techniques have been developed to change the content of images in the sense of image composition and retargeting.

2.1. Image composition and aesthetics

Composition is the arrangement of visual elements in the image frame, which is an essential aspect in the creation of a vast variety of artistic work. In their daily work, professional photographers bring to bear a wealth of photo composition knowledge and techniques [MB98]. No absolute rules exist that ensure good composition in every photograph; rather, there are only some heuristic principles that provide a means of achieving an eye-pleasing composition when applied properly. Some of these principles include: rule of thirds, shapes and lines, amputation avoidance, visual balance, and diagonal dominance [Kra05].

There has been several attempts to allow automatic images cropping or capturing based on the visual quality of the output. Simple techniques from traditional artistic composition have been applied to the artistic rendering of interactive 3D scenes [KHRO01]. The work of Suh et al. [SLBJ03] develop a set of fully automated image cropping techniques using a visual salience model based on low-level contrast measures [IKN98] and an image-based face detection system. [GRMS01] uses the rules of thirds and fifths to place silhouette edges of 3D models in view selection. [BDSG04]

positions the features of interest in an automatic robot camera using the rule of thirds. [LFN04] considers some balance heuristic to arrange images and text objects in a window. Zhang et al. [ZZS*05] propose 14 templates that utilize composition rules to crop photos by using face detection results. Santella et al. [SAD*06] present an interactive method based on eye tracking for cropping photographs. Instead of improving aesthetics, Wang and Cohen [WC06] propose an algorithm for composing foreground elements onto a new background by integrating matting and compositing into a single optimization process. Recently, a quality classifier that assesses the composition quality of images is statistically built using large photo collections available on websites [NOSS09]. The cropped region with the highest quality score is then found by applying the quality classifier to the cropping candidates.

Other attempts to improve image aesthetics modify aspects other than image composition. For example, Cohen-Or et al. [COSG*06], seek to enhance the harmony among the colors of a given image; Leyvand et al. [LCODL08] enhance the attractiveness of digital faces based on a training set.

2.2. Image retargeting

Image retargeting deals with displaying images on small screens such as cell phone displays. The goal of retargeting is to provide effective small images by preserving the recognizability of important image features during downsizing.

Setlur et al. [STR*05] segment an image into regions and identifies important regions. Then, important regions are cut and pasted on the resized background, where missing background regions are filled using inpainting. In our work, we extract salient regions similartly, and use them as primitives in the aesthetic objective function.

The relative distance and distributions of salient objects around the image play a crucial rule in its aesthetics. We therefore employ non-homogenous warping techniques to alter the compositions of the given images. One of the first systems to allow such warpings subject to region-preserving constraints was by Gal et al. [GSCO06], who present a mapping that preserves the shape of important features by constraining their deformation to be a similarity transformation.

Avidan and Shamir [AS07] propose a content-aware approach where a seam-carving operator changes the size of an image by gracefully carving-out pixels in unimportant parts of the image. The seam-carving operator is extended to video retargeting and media retargeting [RSA08, RSA09]. The work of Wolf et al. [WGCO07] presents a retargeting solution for video, in which the warping is computed as a solution for a linear set of equations. Wang et al. [WTSL08] propose an optimized scale-and-stretch approach for resizing images. Recently, some patch based methods are proposed to edit images by allowing modifications of the relative position of objects [CBAF08, SCSI08, BSFG09].

Restricted to still images, the work of Wolf et al. pro-

poses an alternative to the work of Avidan and Shamir. While both methods are efficient and effective, we choose to use the method of Wolf et al. since it seems to produce less artifacts due to its continuous nature. Similarly to Avidan and Shamir's Seam Carving method, the method of Wolf et al. [WGCO07] takes as input a saliency map F and a new image width W_{new}. The treatment of vertical warping is done independently and in an analog manner. The method then solves a system of equations where the new location $x_{i,j}$ of each pixel (i,j) along the x axis is an unknown. The location of the leftmost column of pixels in the new image is set to be 1, and the rightmost column is constrained to be W_{new}. Two types of equations are used to constrain the remaining pixels:

$$F_{i,j}(x_{i,j} - x_{i-1,j}) = F_{i,j} \qquad (1)$$
$$W(x_{i,j} - x_{i,j+1}) = 0 \qquad (2)$$

The first type of equations encourages pixels to be warped at a distance of one pixel apart from their neighbors to the left, and the second type encourages pixels to be warped by the same amount of their neighboring pixel below. The system of equations is solved in a least squares manner, and according to the saliency map F and the weight W, some of the constraints get priority over others. In particular, salient pixels keep their space, while less salient pixels are "squished". The end result is a warping which is smooth, and which more often than not produces images that seem natural.

3. Overview

Increasing the aesthetics of a given image is a twofold problem: how to modify the image and how to measure its new aesthetics. The answer to the latter question is the core of our method. In Section 4 we describe the specific image properties we measure, and how these are computed algorithmically.

As for the first problem, our method employs a compound operator as means to modify a given image: it *non-homogeneously* retargets a cropped part of the image into a target frame having different dimensions than the original image. Then the results are remapped *homogeneously* to the dimensions of the original image. This multi-stage operator modifies the proportion, the interrelation among the geometric entities, and the composition of the image.

The parameters of the above recomposition operator constitute a 6D space. The cropping frame has four degrees of freedom and the target frame two. To reduce the dimensionality of the search space, we limit the crop and target frames to have the same aspect-ratio as the input image, reducing the number of parameters to four: x and y position of the cropping frame, its width, and the amount of retargeting, see Figure 2.

To further reduce the search space, we limit the size of the crop and target frames to be no less than 75% of the original frame size. In Section 5 we show that this reduced search space is effective enough to improve the aesthetics of a given image without causing a dramatic change to the semantics of the original image.
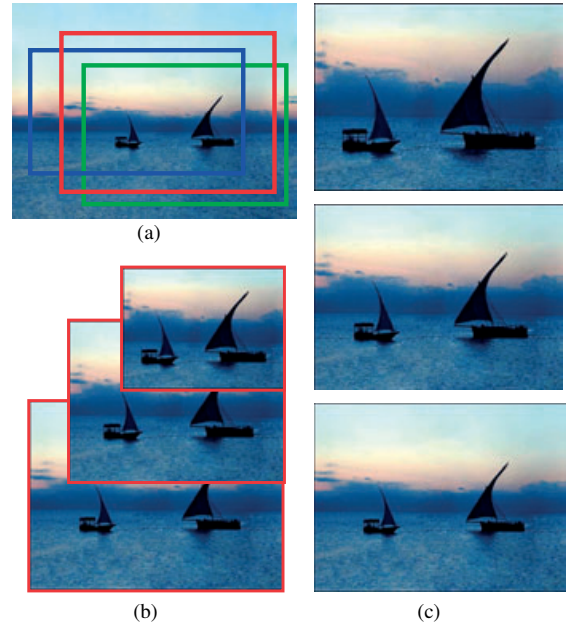

(a)

(b)   (c)

Figure 2: *Overview of our aesthetic retargeting method. (a) The original image with different cropping frames; (b) The red cropping frame in (a) is retargeted into three different frames of the same aspect ratio; (c) The retargeted images in (b) are uniformly scaled to frames of the original sizes, in order to allow a direct comparison between images. Note that the sizes of salient objects and the distances between them are changed by the retargeting operator. The topmost image in (c) displays the most aesthetic result found.*

4. Aesthetic measurement

Our approach is based on searching, in a low-dimensional parameter space, for the most aesthetic image. This is made possible through a computational model of image aesthetics, which bridges between low- and mid-level image primitives and high-level professional guidelines that are often followed.

4.1. Basic aesthetic guidelines

There are various guidelines for shooting well-composed photographs. We consider a limited set of such guidelines that are well-defined and prominent in many aesthetic images.

Rule of thirds The most familiar photo composition guideline is the rule of thirds [GS90, Kra05]. The rule considers the image to be divided into 9 equal parts by two equally spaced horizontal lines and two such vertical lines, as in Figure 3(a). The four intersections formed by these lines are referred to as "power points", and photographers are encouraged to place the main subjects around these points, and not, for example, at the center of the image. Also by this composition-rule, strong vertical and horizontal components or lines in the image should be aligned with those lines. Figure 3(a),(b) demonstrate two aesthetic photographs that comply with this rule.
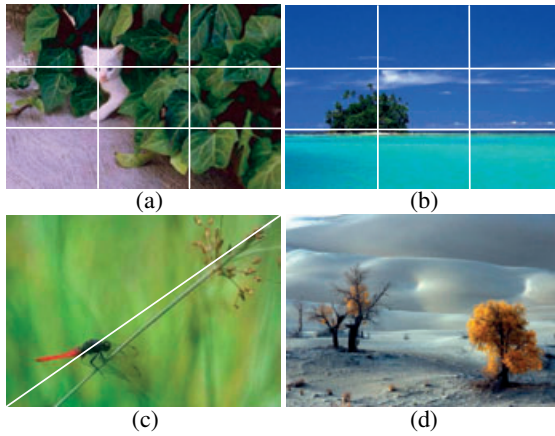
Figure 3: *Basic composition guidelines and examples. (a) the cat object is located at one of the "power points", the thirds lines are overlayed for illustration; (b) the horizon is located at the thirds line; (c) a dominant diagonal component; (d) a balanced image: objects are evenly spread around the center.*

Diagonal dominance In addition to the lines that mark the thirds, the diagonals of the image are also aesthetically significant. A salient diagonal element creates a dynamic emphasizing effect [GS90]. Indeed, one of the most common and effective uses for the diagonal is as a leading line – a line that causes the eyes of the viewers to fixate on the subjects along it. Figure 3(c) shows one such example.

Visual balance The concept of balance is a crucial component to the harmony of an image-composition [Kra05]. In a visually balanced image, the visually salient objects are distributed evenly around the center Figure 3(d). Similarly to a balanced weighing scale, when balanced, the center of the "visual mass" is nearby the center of the image, where this mass-analog takes into account both the area and the degree of saliency of visually salient regions.

4.2. Image pre-processing

The aesthetic score that we assign to an image is based on an analysis of its spatial structure and the distributions of salient regions and prominent lines in the image. The salient region detection is performed using conventional algorithms.

Detecting salient regions The salient regions are detected in a similar manner to what was done in the retargeting system of Setlur et al. [STR*05]. First, we segment the image to homogenous patches using a graph-based segmentation technique [FH04]. We then assign a saliency value to each image-pixel based on a low-level saliency score of Itti et al. [IKN98]. The saliency score is then assigned for each patch by averaging the saliency of the pixels that it covers. Salient patches are then expanded using a greedy algorithm [STR*05] that incorporates nearby patches that share similar color histograms to produce larger salient regions.
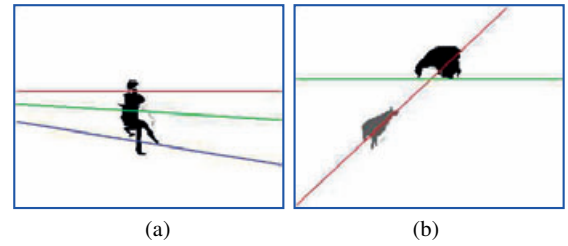

Figure 4: *Detection of salient regions and prominent lines in images (a) Figure 1(a); (b) Figure 7(Lower left). The red line has higher saliency value than the green and blue ones. The darker the regions are, the larger the salience value are.*

Detection of prominent lines Our line detector follows many similar algorithms. First, all line segments along region boundaries in the segmentation result are collected. The boundaries are split by fitting a sequence of straight line segments. Then, out of the infinite straight lines that contain the line segments, the one straight line with the largest support is selected. This selected line is refined based on the participating segments, and trimmed according to the support. The supporting segments are removed, and the process repeats.

In addition to the line detector, we also fit lines to elongated salient regions that may exist in the image. For each detected salient regions S_i in the image, we examine the covariance matrix of the coordinates of all its pixels. If the ratio of the first and the second eigenvalue of this 2×2 matrix is larger than a threshold ($\theta_r = 3$), we fit a line segment to the pixels of the region S_i. This line segment is added to the list of detected lines, and all pixels from S_i that lie on this segment are considered its support. Each detected line L is assigned a saliency value $I(L) = (s_1 + s_2 + s_3)/3$, where s_1 is the total length of the projections of all line segments that support L, s_2 is proportional to the length of L, s_3 is the median value of the norm of the gradient (computed by the Sobel operator) of the pixels along the line L, and all three values are normalized to be no more than one. The higher the value of $I(L)$ is, the more important the prominent line L is in the image. Those with very low saliency values are discarded. Figure 4 depicts examples of salient regions and prominent line detections.

4.3. Aesthetic measurement computation

Given the salient regions, prominent lines, and the computed saliency map, we define a score that evaluate the aesthetics of the image based on the four above-mentioned criteria.

The symbols used in our paper are listed in Table 1. The set X of approximately diagonal lines contains the indices of all lines that form a similar angle with the horizon or the vertical as either Q_1 or Q_2 (we use a threshold of 10 degrees). $\overline{X}$ denotes the set of all other lines. $I(S_i)$ and $I(L_i)$ are explained in Section 4.2.

The normalized Manhattan distance d_M is used to measure distances between 2D points in our system. It is defined

Symbol	Meaning
w, h	The width and height of the image
C	Center of the image frame
$R_i, i = 1, 2, 3, 4$	Four third lines of the frame
$G_i, i = 1, 2, 3, 4$	Four power points of the frame
Q_1, Q_2	Two diagonal lines of the frame
$S_i, i = 1, 2, \cdots, n$	Salient regions detected in the image
$C(S_i), A(S_i), I(S_i)$	Center, area and saliency value of region S_i
$r(S_i)$	Region size – area ratio of S_i with respect to the image
$M(S_i) = A(S_i)I(S_i)$	"Mass" of salient region S_i
$L_i, i = 1, 2, \cdots, m$	Prominent lines detected in the image
X	Indices of approximately diagonal image line
$\overline{X}$	Indices of non-diagonal lines
$I(L_i)$	Saliency value of prominent line L_i
d_M	Normalized Manhattan distance
d_{LM}	Mean points on line distance to line

Table 1: *Symbols used in the paper.*

as $d_M((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| / w + |y_1 - y_2| / h$, where $d_L(L, M)$, the distance measure between two line segments L and M, is defined as the average d_M distance between all points on the segment L and the closest points on M. Since the Manhattan distance is used, the closest point tends to the horizonal or vertical projection, and a closed form formula is easily obtained.

Rule of thirds (RT) The score of this rule has two parts:

$$E_{RT} = \gamma_{point} E_{point} + \gamma_{line} E_{line} \tag{3}$$

where the point score E_{point} measures how close the salient regions lie to the power points, E_{line} measures how close the prominent lines lie to the third lines, $\gamma_{point}, \gamma_{line}$ are weights.

The point score of all salient regions is calculated as:

$$E_{point} = \frac{1}{\sum_i M(S_i)} \sum_i M(S_i) e^{-\frac{D^2(S_i)}{2\sigma_1}} \tag{4}$$

where $D(S_i) = \min_{j=1,2,3,4} d_M(C(S_i), G_j)$ is the minimal distance from the subject center to the four power points G_j, and $\sigma_1 = 0.17$.

The line score is calculated as:

$$E_{line} = \frac{1}{\sum_{i \in \overline{X}} I(L_i)} \sum_{i \in \overline{X}} I(L_i) e^{-\frac{D_R^2(L_i)}{2\sigma_2}} \tag{5}$$

where $D_R(L_i) = \min_{j=1,2,3,4} d_L(L_i, R_j)$ is the minimum line distance between L_i and the third lines, and $\sigma_2 = 0.17$.

In our experience the line based rule of thirds is a better aesthetic predictor than its point-based counterpart and we set the weights in Eq. 3 above as $\gamma_{point} = \frac{1}{3}, \gamma_{line} = \frac{2}{3}$.

Diagonal dominance (DA) The diagonal dominance score is computed similarly to the line based rule of thirds above:

$$E_{DA} = \frac{1}{\sum_{i \in X} I(L_i)} \sum_{i \in X} I(L_i) e^{-\frac{D_Q^2(L_i)}{2\sigma_2}} \tag{6}$$

where $D_Q(L_i) = \min(d_L(L_i, Q_1), d_L(L_i, Q_2))$.

Visual balance (VB) An arrangement is considered balanced if the "center of mass" which incorporates all salient regions is
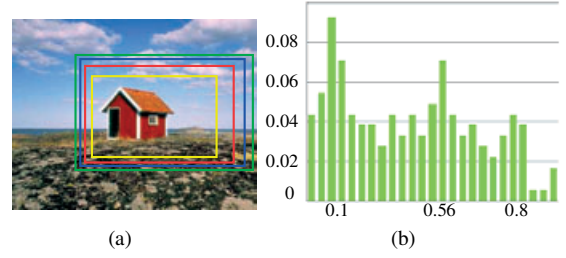

(a)   (b)

Figure 5: *Salient-regions sizes. (a) All the cropping frames have the same maximal scores of E_a if the house object is placed on the power-points of the frames. (b) The histogram of the sizes of salient regions in a versatile set of over 200 professional images.*

nearby the image center C. The visual balance score is therefore ($\sigma_3 = 0.2$):

$$E_{VB} = e^{-\frac{d_{VB}^2}{2\sigma_3}} \tag{7}$$

where $d_{VB} = d_M \left(C, \frac{1}{\sum_i M(S_i)} \sum_i M(S_i) C(S_i) \right)$.

Aesthetic score function (RZ) The aesthetic score function is defined as a combination of the above aesthetic measurement scores:

$$E_a = \frac{\omega_{RT} E_{RT} + \omega_{DA} E_{DA} + \omega_{VB} E_{VB}}{\omega_{RT} + \omega_{DA} + \omega_{VB}} \tag{8}$$

where $\omega_{RT} = 1$ and $\omega_{VB} = 0.3$ are fixed weights. ω_{DA} is 1 if there are detected diagonal lines in the image, zero otherwise.

Salient-regions sizes While combining the three aesthetic guidelines is superior to using just one rule (e.g., the rule of thirds), it turns out that this combined score is not restrictive enough. Considering a simple example that contains only one salient object, this object can be placed on the power-points of the image (rule of thirds) at any scale, see Figure 5(a). That is, there are many cropping frames that have equal highest scores. We now introduce the region size score that plays an important rule in stabilizing the optimization problem by eliminating much of this freedom.

The region size score's main function is to determine the most visually appealing scale. It is based on an observation that region sizes in professional photographs are distributed unevenly. Refer to Figure 5(b), which shows the histogram of the sizes of automatically detected salient regions in a database of more than 200 professional images we collected for this study. Although the images were taken from various sources, and the set of images is very diverse, the underlying distribution is three-modal, and has three dominant peaks that correspond to small regions, medium sized regions, and large regions. In our search for the most pleasing retargeted image, we encourage region of sizes that adhere to this distribution.

Let $r(S_i)$ be the fraction of the image size S_i captures. The sizes of salient regions in aesthetic images are mostly distributed around the values: $r_1 = 0.1, r_2 = 0.56$, and $r_3 = 0.82$,

Image	Sum	RT	DA	VB	SZ
(a)	0.85	0.62	0.00	0.10	0.13
(b)	0.86	0.64	0.00	0.09	0.13
(c)	0.90	0.32	0.36	0.12	0.10
(d)	0.93	0.61	0.00	0.17	0.13

Table 2: *The aesthetic scores for the images in Figure 3.*

corresponding to small, medium and large regions. The size score encourages regions to distribute similarly:

$$E_{SZ} = \sum_i \max_{j=1,2,3} e^{-\frac{(r(S_i)-r_j)^2}{2\tau_j}} \qquad (9)$$

where $\tau_1 = 0.07, \tau_2 = 0.2, \tau_3 = 0.16$ were evaluated by fitting a mixture of Gaussians to the histogram of Figure 5(b).

Combined aesthetic score function The combined score function is defined as a combination of E_a and E_{SZ}:

$$E = (1 - \omega_{SZ})E_a + \omega_{SZ}E_{SZ} \qquad (10)$$

where $\omega_{SZ} = 0.08$. All the weights used in the score function are chosen empirically on a separate set of images, and are fixed for all experiments.

We use our aesthetic score function to calculate the scores of the images in Figure 3. The scores are shown in Table 2. Here, and in the diagrams throughout this paper, the values RT(rule of thirds), DA(diagonal dominance), VB(visual balance) and SZ(region size) correspond to the energy functions (E_{RT}, E_{DA}, E_{VB} and E_{SZ}) weighed as in Eq. 10.

4.4. Optimization

The cropping frames in the original image are searched over a 3D space which consists of the location (x, y) and the width w of the composition rectangle, keeping the aspect ratio of the original image. Then, the cropping frames are retargeted into the target frames by the non-homogenous warping technique [WGCO07], where the amount of retargeting in both axes constitutes a fourth parameter. Figure 6 illustrates how the various aesthetic scores change as a function of one of these four parameters.

The optimization process consists on finding in the 4D parameter space the parameter vector that maximizes the aesthetic score given in Eq. 10. In our system, we seek the optimal solution using particle swarm optimization (PSO) [KE95]. PSO is an evolutionary optimization method starting from many random initialization seeds, where at each iteration a set of solutions exist, the scores of each solution is calculated, and the solutions are updated by shifting them toward the maximal current solution.

5. Results, Validation and Discussion

Figures 7 and 8 show examples of aesthetic composition. Please refer to the supplementary material and video for additional results.

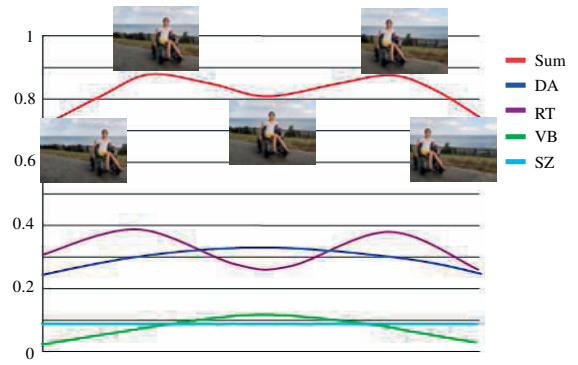The visual balance contributes much to the improvement

Figure 6: *The change in the objective function as the crop window moves from left to right in the image of Figure 1(a). The x-axis depicts the shift in the window location, and the y-axis the resulting score. For this visualization, the y coordinate and the width of the cropping window are fixed, as is the amount of retargeting.*

in Figures 8(a) and (d). The rule of thirds and the diagonal rule are, as expected, anticorrelated. This is much more so in the output images than in the input images. Figure 8(c) places a strong linear-element along the main diagonal. The remapping of Figure 8(b) increases the region size term of the aesthetic score considerably. Note that the relative distances among the objects are modified due to the warping technique in the search, as is very notable in Figure 8(d).

Figure 1 shows another example. There is one prominent horizontal line and two diagonal lines in the original image, see Figure 1(a). Optimizing this image leads to the new recomposed image (Figure 1(c)) that obtains a higher aesthetic score than any cropping frame such as the one shown in Figure 1(b). It is observed that the result of Figure 1(c) is not just a cropping of Figure 1(a) as it contains much more cloud than the corresponding cropping frame.

The proposed set of aesthetic rules work in unison in the score function, see Figure 9. The rule of thirds alone, which dominates previous work, is not enough for ensuring appealing composition. A statistical analysis reveals that due to the high weight assigned to it, the rule of thirds, applied both to points and lines, dominates the total score in the original image, however at the output image, the contribution is more evenly spread among the various aesthetic guidelines. This is the case in both examples in Figure 7. Also, in the original images, visual balance and the rule of thirds are uncorrelated. In the output images, they become highly correlated. Examples of the interplay between the various rules can be observed by examining the bar plots of Figures 1(d), 7(d), and 12(d), and the graphs of Figure 6.

To numerically evaluate our score function, we employed a dataset of 900 casual images arbitrarily collected from international websites in which skilled photographers rank photographs through them: 300 of the top-ranked images, 300
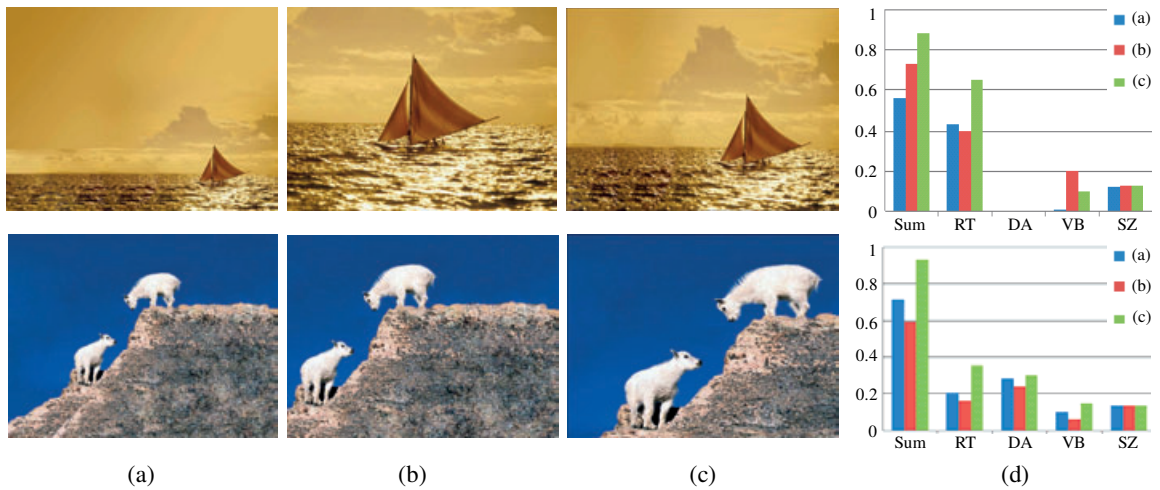
Figure 7: *Results of aesthetic composition. (a) The original images; (b) an arbitrary cropping frame of (a); (c) the aesthetic composition result by our approach; (d) the aesthetic scores of (a),(b),and (c).*
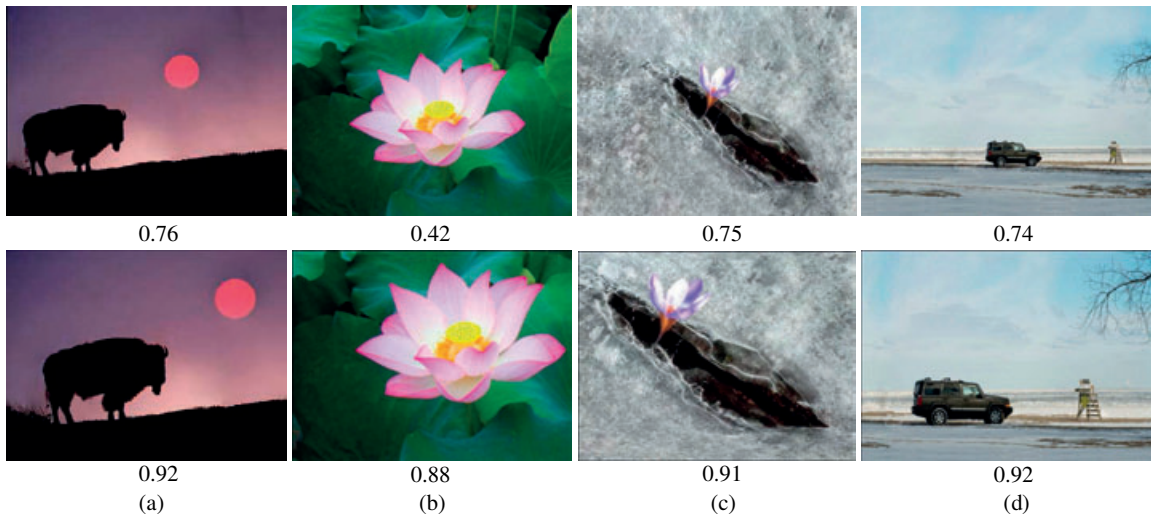

Figure 8: *More results generated by our method. Upper row: original; Lower: optimized. The salient regions in (c) and (d) are detected in a semi-automatic fashion. The numbers indicate the aesthetic scores.*

ranked as good, and 300 casual images were collected. We compute the aesthetic scores for these photos and their optimized versions. The histograms are shown in Figure 10. As can be seen, the aesthetic score we devise is spread differently among the three groups, and all three histograms move to the region of high scores during the optimization process.

To further study our methods, we have compared it to existing recomposition methods [SLBJ03] and [SAD*06]. Instead of using the eye tracking data, we use the same salience map to run the algorithm of [SAD*06] as used in the other approaches. Note that these methods have been designed to maximize other scores: [SLBJ03] maximizes the crop's saliency, and [SAD*06] maximizes content area and features.

Also note that these methods are confined to simple cropping. As can be seen in Figure 11, the method of [SAD*06] does not produce particularly aesthetic results. The method of [SLBJ03] produces somewhat simpler images, and aims to create thumbnail images that are easily recognizable. To prevent a bias in the results due to selection of the input images, we made sure to include many casual images and the images of the Berkeley Segmentation Benchmark [MFTM01] in our experiments. The results are provided in the supplementary material. While no method can recover from a very poor input composition, a good system is expected to either create a noticeably better composition or to keep the input composition more or less the same. As shown in the results, our method
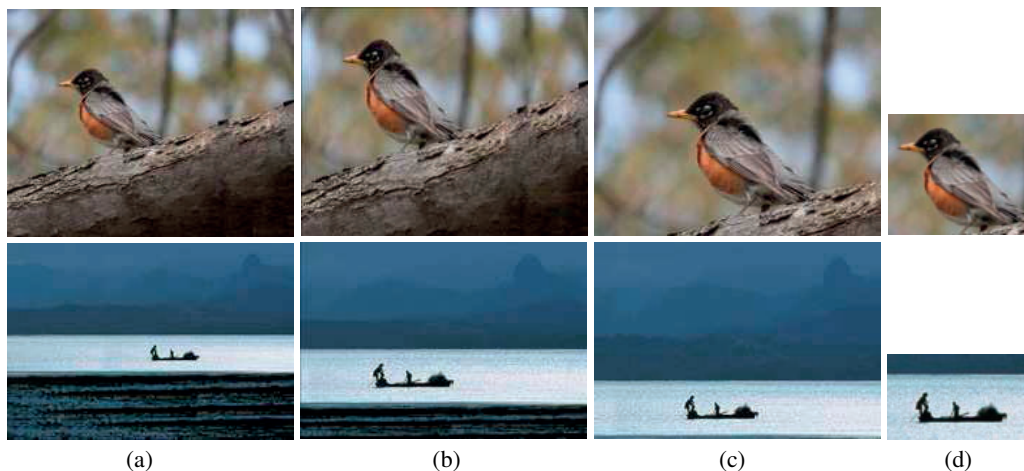
Figure 11: *Comparison with the previous approaches. (a) The original images; (b) the results of our approach; (c) the results of Santella et al.'s approach [2006]; (d) the results of Suh et al.'s [2003]. Note that line-based information plays a crucial role in photo composition and ignoring it leads to inferior results.*

Figure 9: *Aesthetic composition results using different guidelines. Upper row: (a) the original image; (b) the composition result using the guideline of rule of thirds; (c) the composition result using the guideline of rule of thirds and diagonal dominance. Lower row: (a) the original image; (b) the composition result using the guideline of rule of thirds; (c) the composition result using the guideline of rule of thirds and visual balance. Here $\omega_{SZ} = 0$ to see how the other composition rules work.*

Figure 10: *(a) Aesthetic score histograms for three sets of photographs: red bars show the results of professional-level photos; green – good photos; blue – casual photos. (b) Same, for the matching optimized images.*

is robust in that sense (see the comparison with [SAD*06] in the supplementary material).

The crop-and-retarget operator typically results in a zoom-in effect. For some images, however, the most aesthetic result is obtained by capturing a larger zoomed-out frame. If the background is simple, we can use texture synthesis or in-painting techniques to enlarge the image prior to applying our technique. Figure 12 contains one such example, where the image was extended using a texture synthesis technique. Indeed, the most aesthetic version of this example has a frame larger than the original one.

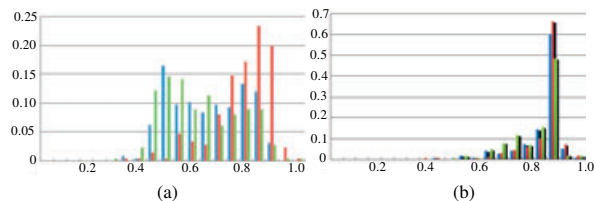The same "zoom-out" technique can also be used to objectively validate our performance. We have collected a test set of professional photographs and extended their content by means of texture-synthesis. We then applied our method to the photographs. As can be seen in Figure 13, the recomposed photographs look similar to the original photographs.

Our algorithm takes 0.14-0.18 sec to optimize the composition a photo of size 1024×768 if we only allow cropping in the searching. If we incorporate retargeting operator, it takes 2-14 sec.

User study To further evaluate the performance of our method, we have conducted three user studies. The first compared viewers' assessment of the aesthetic appeal of our approach and gaze-based cropping approach [SAD*06]. We have generated a set of 30 triples of images; one original, one crop generated by [SAD*06], and one generated by our approach. Each subject was asked to select the best looking image out of each triple. The second user study involved examining whether our optimized results are competitive to the crops by a professional photographer. For a set of 30 images, the skilled photographer cropped a "best looking" crops for each image by hand in Photoshop. The optimized images were generated using our approach. Each subject was asked to
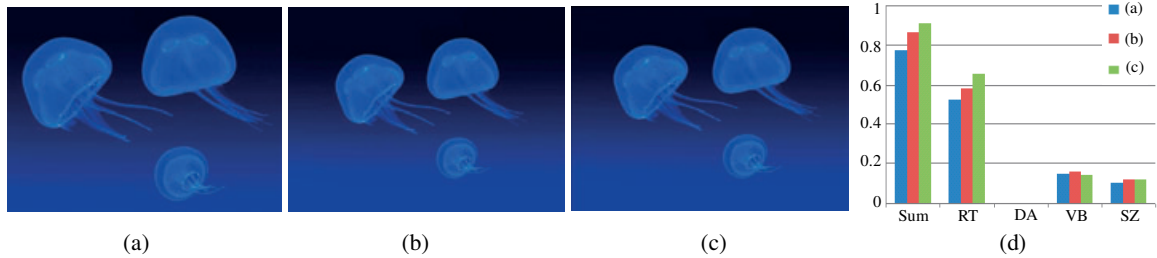
(a)   (b)   (c)   (d)

Figure 12: *For some images, the best aesthetic results are obtained by zooming out. While zooming out is not possible for all images, for some images texture-synthesis enables such an effect. (a) The original image; (b) Enlarged image by applying a texture synthesis technique; (c) the result generated by applying our technique on (b); (d) the aesthetic scores of (a),(b), and (c), respectively.*

Figure 13: *Upper left: the original professional photographs; Bottom: the extended images by Photoshop; Upper right: optimized images from the extended images. Note that the recomposed images are typically similar to the original images, which means that our method is able to identify the the aesthetic quality of the original image.*

select whether one image looks much better than the other or whether "the two images look similar". The third user study aimed to assess the performance of our method. This time, the subjects were first taught some basic composition guidelines as shown in Section 4.1. Once again, 30 pairs of original image and optimized image were shown to each subject who was asked which better adheres to the guidelines. In all the studies, the test images are randomly chosen and the images in each pair or triple are shown side by side (random order) on a 19-inch CRT. A total of 56 subjects each participated in the three sets of experiments, which took about 20 minutes on average. The subjects are males and females between the ages of 21 and 55. 7 subjects have much photography or art experience, 33 subjects have a few knowledge of photography, and the others almost know nothing about photography.

In the first study, the percentages of the selected original images, the results of gaze-based method, and the results of our method are $19.6\%, 36.3\%$ and 44.1%, respectively. It shows a clear tendency toward the recomposed images using our approach.

In the second study, the percentages of the selected answers of "hand-cropped image", "similar", and "optimized image" are $15.2\%, 81.8\%$ and 3.0%, respectively. It shows that the

optimized images generated by our approach are close to the professional crops.

In the third study, the users agree almost unanimously (93.7%) that the manipulated images better adhere to the given composition rules.

Limitations Professional photographs do not necessarily use the predefined aesthetic guidelines, and often chose to disobey them. Our technique follows the guidelines without discretion and does not apply inspiration or creativity. Moreover, our method, similarly to any method that modifies the relative locations of image parts, may change relative sizes and proportions within the image such that the image semantics are altered. For some images, the salient regions detection algorithm does not detect all salient regions. We therefore applied this algorithm in a semi-automatic fashion and augmented the list of salient regions. As the images are warped in the retargeting, distortion on the salient objects might be noticeable in some results.

6. Conclusion and Future Work

We demonstrate that aesthetics can be evaluated computationally with high enough accuracy to be useful. This opens a new avenue for various applications to be enhanced by the ability to automatically assign aesthetic scores. For example, aesthetic views of 3D models can be identified and appealing logos can be generated given a set of user requirements.

As a first such application we propose the ability to automatically recompose images, and show that by optimizing a set of only four parameters we are able to generate recomposed images that are notably more aesthetic. Future efforts for the recomposing application, can focus on improving the aesthetic score. We would like to explore the possibility of improving the salient-region detection method by means of computational-learning, and to add color based considerations to the score enabling the automatic augmentation of the colors in the image. Aesthetics perception is also influenced by the structure of the underlying scene, and we would like to explore adding this and semantic information to the analysis.

References

[AS07] AVIDAN S., SHAMIR A.: Seam carving for content-aware image resizing. *ACM Transactions on Graphics 26*, 3 (2007). 2

[BDSG04] BYERS Z., DIXON M., SMART W. D., GRIMM C.: Say cheese! experiences with a robot photographer. *AI Magazine 25*, 3 (2004), 37–46. 2

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D. B.: Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics 28*, 3 (2009). 2

[CBAF08] CHO T. S., BUTMAN M., AVIDAN S., FREEMAN W. T.: The patch transform and its applications to image editing. In *Proceedings of CVPR* (2008). 2

[COSG*06] COHEN-OR D., SORKINE O., GAL R., LEYVAND T., XU Y.-Q.: Color harmonization. *ACM Transactions on Graphics 25*, 3 (2006), 624–630. 2

[FH04] FELZENSZWALB P. F., HUTTENLOCHER D. P.: Efficient graph-based image segmentation. *International Journal of Computer Vision 59*, 2 (2004), 167–181. 4

[GRMS01] GOOCH B., REINHARD E., MOULDING C., SHIRLEY P.: Artistic composition for image creation. In *Proc. of the 12th Eurographics workshop on Rendering Technique* (2001), pp. 83–88. 2

[GS90] GRILL T., SCANLON M.: *Photographic Composition*. Watson-Guptill, 1990. 1, 3, 4

[GSCO06] GAL R., SORKINE O., COHEN-OR D.: Feature-aware texturing. In *Eurographics Symp. on Rendering* (2006), pp. 297–303. 2

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20* (1998), 1254–1259. 2, 4

[KE95] KENNEDY J., EBERHART R.: Particle swarm optimization. In *Proc. IEEE Conf. on Neural Networks* (1995), pp. 1942–1948. 6

[KHRO01] KOWALSKI M. A., HUGHES J. F., RUBIN C. B., OHYA J.: User-guided composition effects for art-based rendering. In *Proc. of the Symposium on Interactive 3D graphics* (2001), pp. 99–102. 2

[Kra05] KRAGES B.: *Photography: The Art of Composition*. Allworth Press, 2005. 1, 2, 3, 4

[LCODL08] LEYVAND T., COHEN-OR D., DROR G., LISCHINSKI D.: Data-driven enhancement of facial attractiveness. *ACM Trans. Graph. 27*, 3 (2008). 2

[LFN04] LOK S., FEINER S., NGAI G.: Evaluation of visual balance for automated layout. *Intelligent User Interfaces* (2004), 101–8. 2

[MB98] MARTINEZ B., BLOCK J.: *Visual Forces, an Introduction to Design*. Prentice-Hall, New York, 1998. 1, 2

[MFTM01] MARTIN D., FOWLKES C., TAL D., MALIK J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *International Conference on Computer Vision* (2001), vol. 2, pp. 416–423. 7

[NOSS09] NISHIYAMA M., OKABE T., SATO Y., SATO I.: Sensation-based photo cropping. In *Proc. ACM International Conference on Multimedia* (2009). 2

[Pet07] PETERS G.: Aesthetic primitives of images for visualization. In *11th Int. Conf on Information Visualization* (2007), pp. 316–325. 1

[RPJJS07] RIVOTTI V., PROENAA J., J. JORGE J., SOUSA M.: Composition principles for quality depiction and aesthetics. In *The International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging* (2007), pp. 37–44. 1

[RSA08] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Improved seam carving for video retargeting. *ACM Transactions on Graphics 27*, 3 (2008). 2

[RSA09] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Multi-operator media retargeting. *ACM Transactions on Graphics 28*, 3 (2009). 2

[SAD*06] SANTELLA A., AGRAWALA M., DECARLO D., SALESIN D. H., COHEN M. F.: Gaze-based interaction for semi-automatic photo cropping. In *ACM Human Factors in Computing Systems (CHI)* (2006), pp. 771–780. 2, 7, 8

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Proceedings of CVPR* (2008). 2

[SLBJ03] SUH B., LING H., BEDERSON B., JAOBS D.: Automatic thumbnail cropping and it's effectivness. In *ACM Conference on User Interface and Software Technolgy* (2003), pp. 95–104. 2, 7

[STR*05] SETLUR V., TAKAGI S., RASKAR R., GLEICHER M., GOOCH B.: Automatic image retargeting. In *Proc. of Mobile and Ubiquitous Multimedia (MUM)* (2005), pp. 59–68. 2, 4

[WC06] WANG J., COHEN M.: Simultaneous matting and compositing. ACM SIGGRAPH Technical Sketch, July 2006. 2

[WGCO07] WOLF L., GUTTMANN M., COHEN-OR D.: Non-homogeneous content-driven video-retargeting. In *Proceedings of the 11th IEEE International Conference on Computer Vision* (2007). 2, 3, 6

[WTSL08] WANG Y.-S., TAI C.-L., SORKINE O., LEE T.-Y.: Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics 27*, 5 (2008). 2

[ZZS*05] ZHANG M., ZHANG L., SUN Y., FENG L., MA W.-Y.: Auto cropping for digital photographs. In *Proc. of IEEE International Conference on Multimedia and Expo* (2005). 2