A Framework for Photo-Quality Assessment and Enhancement based on Visual Aesthetics

Subhabrata Bhattacharya
University of Central Florida
subh@cs.ucf.edu

Rahul Sukthankar
Intel Labs & Carnegie Mellon
rahuls@cs.cmu.edu

Mubarak Shah
University of Central Florida
shah@cs.ucf.edu

ABSTRACT

We present an interactive application that enables users to improve the visual aesthetics of their digital photographs using spatial recomposition. Unlike earlier work that focuses either on photo quality assessment or interactive tools for photo editing, we enable the user to make informed decisions about improving the composition of a photograph and to implement them in a single framework. Specifically, the user interactively selects a foreground object and the system presents recommendations for where it can be moved in a manner that optimizes a learned aesthetic metric while obeying semantic constraints. For photographic compositions that lack a distinct foreground object, our tool provides the user with cropping or expanding recommendations that improve its aesthetic quality. We learn a support vector regression model for capturing image aesthetics from user data and seek to optimize this metric during recomposition. Rather than prescribing a fully-automated solution, we allow user-guided object segmentation and inpainting to ensure that the final photograph matches the user's criteria. Our approach achieves 86% accuracy in predicting the attractiveness of unrated images, when compared to their respective human rankings. Additionally, 73% of the images recomposited using our tool are ranked more attractive than their original counterparts by human raters.

Category and Subject Descriptors: H.4 [Information Systems Applications] : Miscellaneous

General Terms: Experimentation, Human Factors.

Keywords: Interactive photo tools, spatial recomposition, quality enhancement.

1. INTRODUCTION

According to statistics quoted by Flickr, an average of 6.5 million photographs are uploaded daily by its users. Thus, there is a great demand for multimedia applications to manage, rate and edit such content. Photo-quality assessment and improvement are two areas that have particularly attracted recent research attention.

The notion of a "high quality" image as perceived by a viewer is often an abstract concept, even for professional photographers, which is why assessing the aesthetic quality of photographs is challenging. However, photographs taken by experienced photographers adhere to several rules of composition, which make them more visually appealing than those taken by amateurs. Studies have revealed that such photographic compositions trigger several psycho-visual stimuli in the human observer due to which the photograph is perceived to be of good quality. As described in the photography literature [11], these include the *Rule of Thirds*, and *Visual Weight Balance*. Elementary photography lessons emphasize that adhering to these two rules alone could significantly improve the aesthetic quality of most amateur photographs (see Fig. 1). In order to satisfy the *Rule of Thirds* the photographer places the primary subject of the composition near a location that is a strong focal point. Similarly, according to the rule of *Visual Weight Balance*, in a well composed image the visual weights of different regions satisfy the *Golden Ratio*. We discuss these rules in detail in the subsequent sections.

Research in evaluating photographic quality dates back to the work of Damera-Venkata *et al.* [5] where the authors use a reference image with its noise degraded counterpart to assess its quality. More recently, Ke *et al.* [12] construct high-level features for photo quality assessment extracted from low level cues like noise,

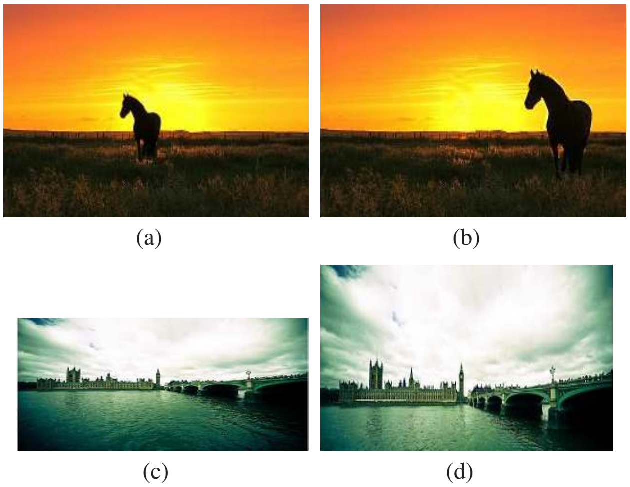


Figure 1: Image enhancement using two independent spatial recomposition techniques proposed in this paper: (a) Original image with a distinct foreground; (b) Aesthetically enhanced image using optimal object placement technique; (c) Original image with unbalanced visual weights; (d) Aesthetically improved image using visual weight balancing technique; Repositioning the horse, and cropping the water-region dramatically enhances the aesthetic appeal of the two photographs.

MM'10, October 25–29, 2010, Firenze, Italy.

blur, color, brightness, contrast and spatial distribution of edges. In addition to some of these low level cues, the authors of [6, 7] investigate the impact of features such as familiarity measures, wavelet responses on textures, aspect ratio and region composition on the aesthetic appeal of natural images. Boutell and Luo [3] explore a variety of metrics including ISO speed rating, F-number and shutter speed, extracted directly from camera metadata, to determine their impact on photographic quality. These methods, as observed by Luo and Tang [16] and Sun *et al.* in [19], capture only fine-grained details about the photograph that are mainly introduced due to sensors used during the image formation process. Thus, in order to understand the nuances of spatial composition in photographic frames, the authors of [16] additionally introduced a parameter that considered adherence to geometric composition rules for photo and video quality evaluation.

While [16] demonstrated some level of success in evaluating photo-quality in natural photographs, that approach relies heavily on a blur detection technique to identify the foreground object's boundary within the frame. This technique works well only with photographs captured using professional SLR cameras that have mechanisms to induce depth-of-field effects and precludes its use with photographs taken using popular point-and-shoot cameras.

We argue that true aesthetic assessment should not be constrained by equipment capability as photographs captured using professional cameras are rarer in number and often restricted by terms of use. Furthermore, photographs captured using professional equipment are more likely to follow composition guidelines since they are generally taken by experienced photographers. Our approach can be perceived as a method to improve photographs, such as those frequently found on the Internet, that were taken by amateurs using consumer digital cameras.

We formulate photo quality evaluation as a machine learning problem in which we map the characteristics of a human-rated photograph in terms of its underlying adherence to the rules of composition. Our method can be compared with the approach suggested in [19, 21], wherein the authors apply a saliency map to estimate visual attention distribution in photographs. We complement the saliency information extracted from an image using a high-level semantic segmentation technique that infers the geometric context [9] of a scene. With the help of the above methods, we extract aesthetic features that could be used to measure the deviation of a typical composition from ideal photographic rules of composition. These aesthetic features are subsequently used as input to two independent Support Vector Regressors in order to learn the visual aesthetic model. This learned model is then integrated into our photo-composition enhancement framework. To this end, we make the following contributions in this paper: (1) Perform an empirical study on visual aesthetics using real human subjects on real-world images, (2) Find a smooth mapping between user input visual attractiveness and high-level aesthetic features, (3) Apply semantic scene constraints while recompositing a photograph, (4) Introduce an interactive tool that helps users to recompose photographs with some informed aesthetic feedback, and finally (5) Bring photographic quality assessment and enhancement under a single unifying framework. An overview of our approach is shown in Fig. 2.

We primarily focus on outdoor photographic compositions with a single foreground object or landscapes and seascapes that lack a dominant object. For the former, we constrain our algorithm to relocate the object to a more aesthetically pleasing location while respecting the scene semantics (e.g., a tree attached to the ground must remain in contact with the ground) and rescaling it as necessary to maintain the scene's perspective. This is a significant improvement over a foreground object-centric image-editing technique [4], wherein the authors propose a method to reconstruct an image from low-resolution patches subject to user-defined constraints. In the case of photographs that lack distinct foreground objects such as land/sea scapes where the dominant portion of the image is covered by sky or sea, we crop or expand the photograph so that an aesthetically pleasing balance between sky and land/sea is achieved. When the spatial alterations create holes where the original photograph lacks information, we apply inpainting to preserve the photo-realism of the original while minimizing artifacts. In this context, our work is partially motivated by the work of Nishiyama *et al.* [18], which introduces a method for automatically cropping a photo using a quality classifier built from user responses; their method implicitly assumes that in a given image, the background region is blurred to emphasize the subject region. Since we rely on a segmentation algorithm that provides us with semantic information of the scene, we can address a broader spectrum of photographs, relaxing this assumption. Our approach is also philosophically similar to Leyvand *et al.*'s work [13] on beautification of human facial images, which quantifies the attractiveness of a human face from the spatial location of features such as eyes, lips and nose, and alters the photograph so as to realign these features to more desirable positions.

We organize this paper as follows. In the next section, we present the details of our approach for learning and assessing the aesthetic quality of a photograph, along with results demonstrating its agreement with human ratings. In Section 3, we discuss our approach to enhance the aesthetic appeal of photographs through the proposed recomposition framework and show examples from our dataset that highlight specific aspects of the process. This is followed by experimental results on assessment and recomposition. Finally, we conclude this paper by discussing several possible applications.

2. LEARNING AESTHETICS

Modern digital cameras employ several auto-focus filters implemented in firmware that also provide an estimate about the focused subject's location in the frame [2]. The photographer (if aware of the rules of composition) could in turn use this real-time aid to adjust the image frame so that the capture conforms to composition guidelines, resulting in an increased aesthetic appeal. However, once a photograph is captured, there is little scope to assess or alter its quality using these camera tools as automatically segmenting the foreground object from an already existing image is a challenging computer vision problem in its own right.

Visual saliency [20] based techniques have been used in the past [19, 21] to obtain a reasonable estimate of the spatial location of the dominant foreground regions in photographs. While this approach addresses our need for identifying the spatial location of the object in a photo-frame, it does not provide any scene semantics that we require to (1) assess the aesthetic appeal based on visual weight, or (2) recompose the given image while maintaining the scene integrity. We tackle this problem using a supervised learning-based scene classification method proposed by Hoiem *et al.* in [9]. This technique generates a confidence map of semantic labels that we can employ to identify likely regions of *sky* and *support* (a generic term for non-foreground regions that do not belong to sky) in an image. Since the images in our dataset are primarily single-subject compositions, the complementary regions in the image that belong to neither the sky nor support, correspond to the foreground (by rule of elimination). We use morphological processing tools to disregard small disconnected regions in order to obtain a reasonable mask for the foreground. Our tool allows users to interactively refine the foreground segmentation and horizon estimation, which is

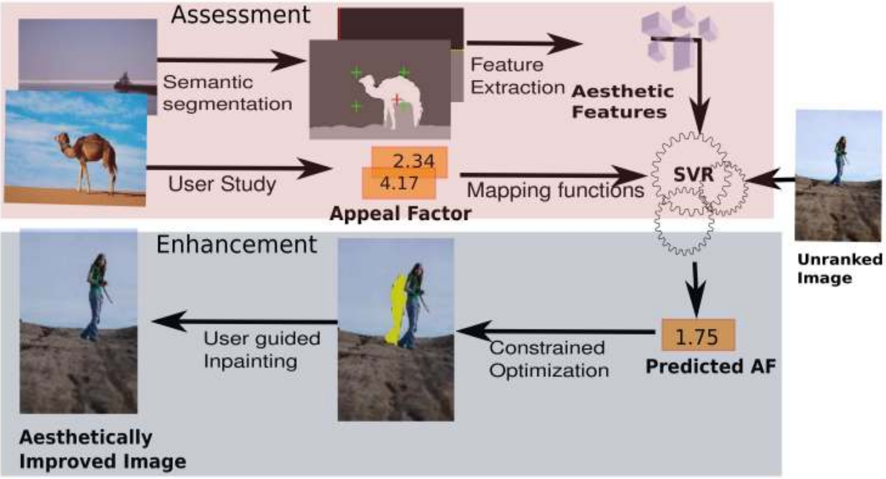


Figure 2: An overview of our photo-quality assessment and enhancement framework. Assessment: we learn a mapping between measured features in an image and the appeal factor using a user study. Enhancement: we generate recompositions that optimize the predicted appeal factor while obeying semantic constraints.

crucial to achieving aesthetically pleasing yet semantically correct results.

2.1 Dataset

Most of the earlier papers [6, 7, 16, 12, 19, 18, 3, 17, 21] on this topic evaluate their respective approaches on their own private collections, which are not made available. In order to contribute to the research community, we make an attempt to build the first dataset of this kind which is reusable, expandable and publicly available. Our dataset[1] consists of 632 digital photographs, all downloaded from free image sharing portals, such as Flickr. Out of these, 384 images conform to the category of single-subject compositions, while the rest are of landscapes or seascapes that do not have any distinct foregrounds. Images that are greater than 640 × 480 in their spatial resolution, are rescaled to this size for computational reasons. Fig. 3 presents a subset of images that we have used in this paper. A Ground Truth aesthetic appeal factor (discussed in Sec. 2.2), associated with each image is used to evaluate the performance of our quality assessment algorithm and is used later to perform the recomposition.

2.2 User Survey

We conducted a thorough study of human aesthetics through a survey where 15 independent participants were asked to assign integer ranks to the photographs in our dataset from 1 to 5, with 5 being assigned to the most appealing. Further, while ranking, users were specifically instructed to eliminate bias from their ratings that might have emerged due to individual subject matter contained in a photograph, e.g., whether a user prefers mountains to sea or birds to animals. Each user was asked to rank no more than 30 images in a particular sitting in order to avoid undesirable variances in the ranking system due to fatigue or boredom. This process was further repeated 5 times to eliminate rankings from inconsistent users. After discarding the scores assigned by inconsistent users, we observed that the distributions were typically unimodal with low variance, enabling us to generate a single ground truth aesthetic appeal factor for each image (F_a) by averaging its assigned scores.

To truly understand how the rules of composition affect the ranking system, on a different setting, participants were divided into 3 groups and a subset of 20 randomly-selected images were assigned to each group. Users of each group were asked to select the foreground and specify a region in background, where they wished the foreground object to be placed while preserving the scene semantics, e.g., the boat stays in water. Perspective correction and images were further touched up to remove segmentation artifacts. The ranking exercise was then interchanged between the groups, and a corresponding F_a is obtained per modified image. We observed the following interesting trends in the rank assignment among the images: (1) images with $1 < F_a \leq 2$ received 91% of the votes marked as 1 and 2, and (2) images with $4 < F_a \leq 5$ received 88% of the votes marked as 4 and 5. This indicates that the participants are clearly able to distinguish between a well-composed and a poorly-composed image based on the foreground's spatial location in the image frame; these results are detailed in Fig. 4.

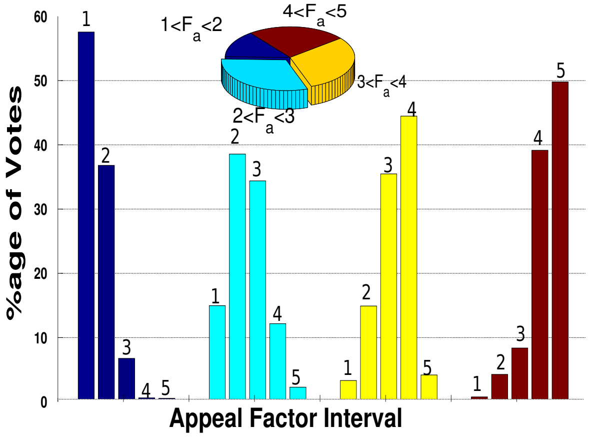
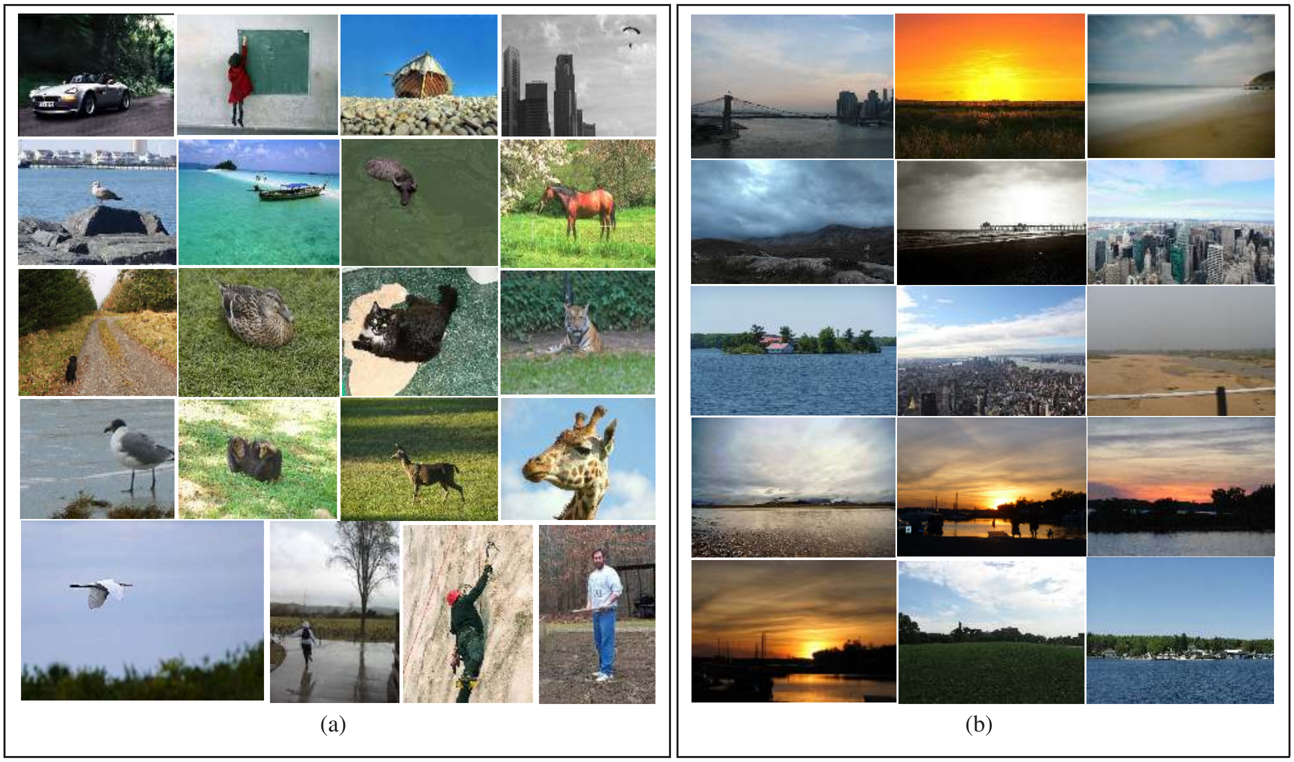


Figure 4: Summarized information from our user survey: The pie-chart on top shows the distribution of the ground truth aesthetic appeal across various images in our dataset. Each sectors in the pie-chart correspond to a discrete appeal factor interval $((1-2], \ldots, (4-5])$. The bar graph in the bottom shows the distribution of the assigned ranks within each interval. For eg. in the interval $1 < F_a \leq 2$, we observe a large number of images that are ranked 1 by most users.

[1] http://www.ucf.edu/~subh/photoquality


(a)

(b)

Figure 3: A small sample of images from our dataset, (a) Images containing single foreground composition, (b) Images of landscapes or seascapes containing no definite foreground.

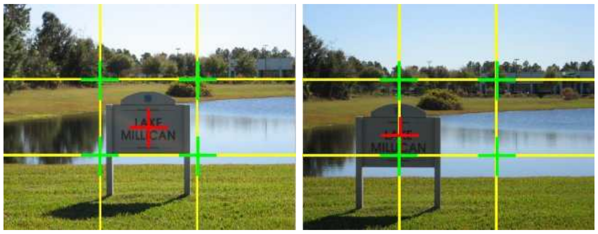
2.3 Aesthetic Features

In order to formulate photographic quality assessment in the context of a machine learning problem, we need to associate the users' notions of aesthetics to well defined, composition-specific features from an image. To this end, we extract a *relative foreground position* feature for images with single-foreground compositions, and a *visual weight ratio* feature for photographs of seascapes or landscapes. Both of these features are based on elementary rules of photographic composition and are discussed as follows:

(a) Relative foreground position is defined as the normalized Euclidean distance between the foreground's center of mass, also called the *visual attention center*, to each of four symmetric *stress points* in the image frame. In photographic literature [11], the stress points are the strongest focal points in a photographic frame(indicated by green cross-hairs in Fig. 5(a)). In order to attract the viewer's attention to a foreground, the photographer is often advised to adjust the frame in a way so that the foreground's center of mass (red cross-hair) coincides with one of these stress points. The clause, "one of these stress points", is of particular interest in this context, since if the visual attention center is positioned equidistant from all the stress points during the capture, the viewers' attention gets equally divided across these four points. This causes the viewer to lose interest in the photograph, thereby reducing its aesthetic appeal (see Fig. 5(a)). This observation is also confirmed by our user study where participants tend to rank images with foreground aligned near a stress point higher than those with foreground centered in the frame.

Thus, every photograph containing a single subject composition can be uniquely characterized by a four dimensional feature vector (**F**):

$$\mathbf{F} = \frac{1}{h \times w} \left[||\mathbf{x}_0 - \mathbf{s}_1||_2, \ldots, ||\mathbf{x}_0 - \mathbf{s}_4||_2 \right], \qquad (1)$$



(a)

(b)

Figure 5: Relationship between Visual attention center and four stress points(adapted from the rule of thirds). Yellow lines divide the rectangular frame into nine identical rectangles. Each intersection of the yellow lines generates a stress point indicated by green cross-hairs, while the red cross-hairs in each image mark the foreground object's visual attention center. (a) A photograph taken with the object placed in the middle of the frame. (b) The same scene photographed after aligning the visual attention center close to the stress-point on the bottom-left. (Best viewed in color.)

where h, w are the height and width of the image, $\mathbf{x}_0$ is the visual attention center and $\mathbf{s}_i$ are the stress points starting from top-left, in clockwise direction. Fig. 6 shows two single subject compositions from our dataset, with their respective visual attention center and stress point locations. Table 1 shows the corresponding appeal factor of these images, obtained from the user study with the computed **F** values. Figs. 6 and 7 demonstrate two automatic techniques that we have used throughout this paper for segmenting the foreground from the background and extracting vital semantic information about the scene.

Although the relative foreground location is effective for typical single- subject compositions, it is inapplicable for the class of

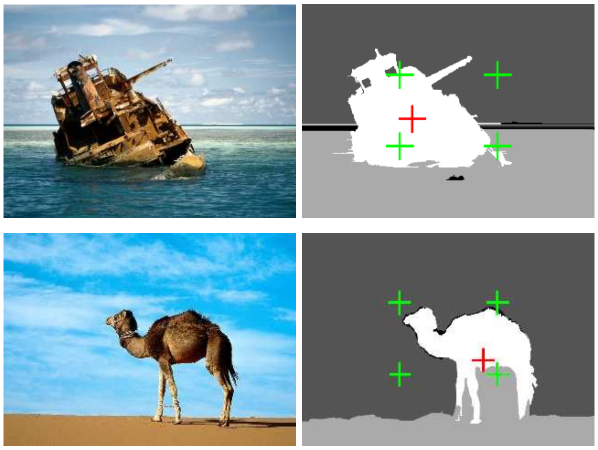


Figure 6: Determining visual attention center using segmentation technique exploiting geometric contexts [9]. Here dark-gray pixels denote sky, light-gray denote support, and white pixels belong to the dominant foreground object. Red and green cross-hairs indicate the locations of the visual attention center ($\mathbf{x}_0$) and the four stress points ($\mathbf{s}_1 \ldots \mathbf{s}_4$) in the frame. Note that the foreground object's outline is more detailed in this case, compared to the saliency based technique illustrated in 7, which makes the former a better fit for the recomposition technique, discussed later.The adjacent table in 1 shows a mapping between the aesthetic appeal factor (F_a) and the relative foreground location feature (F), extracted from these two images.

$\mathbf{F_a}$	Relative Foreground Location (**F**)			
	Top-left	Top-Right	Bottom-Right	Bottom-Left
4.25	0.2940	0.4451	0.3365	0.0399
4.17	0.4381	0.4477	0.0233	0.3935

Table 1: Mapping user rated Appeal Factor (F_a) to the Relative Foreground Location feature (F) for ship and camel images in Fig. 6. The values in the second to fifth columns can be interpreted as the relative Euclidean distances between the visual attention center($\mathbf{x}_0$) and the four stress-points ($\mathbf{s}_1 \ldots \mathbf{s}_4$), normalized against the width and height of the image frame.

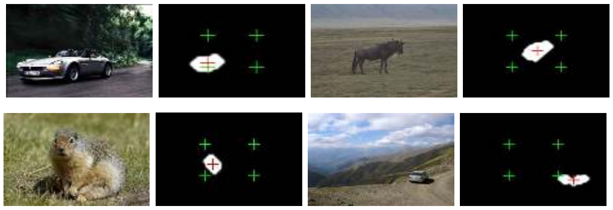
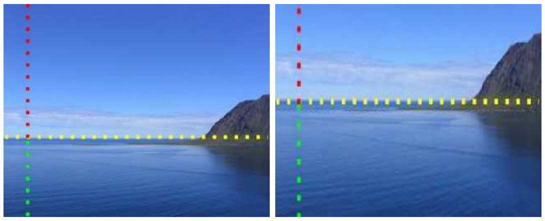


Figure 7: Saliency based detection of visual attention center. Each row shows two pairs of input and output images, the images in black background rows show the output of the saliency algorithm proposed in [20]. Dominant foreground region is shown as a white blob in a black background. Similar to Fig. 6, the visual attention center and the four stress points are shown in red and green cross hairs respectively. Note this technique by itself does not provide us with any scene information.

images in our dataset that consist of landscape or seascape scenes, lacking a compact foreground object. For such images, we formulate a second set of features.

(b) Visual weight ratio can be described as the ratio of approximate number of pixels in the sky region, to that in the support region (ground or sea). We estimate the visual weights in the sky region by the automatic semantic segmentation technique discussed in the beginning of this section. Our tool allows the user to interactively adjust the detected horizon line.



(a) (b)

Figure 8: Quantifying visual weight balance: the yellow dotted line marks horizon, the red dotted line marks the vertical extent of sky (Y_k), and the green dotted line marks the vertical extent of the sea (Y_g). (a) Composition with ideal combination of visual weights. (b) Cropped version of the same composition with altered visual weights. The former is rated as more visually appealing.

The idea behind visual weights can be illustrated with the help of Figs 8(a) and 8(b). In both of the images, the horizon separates the frame into two rough rectangles. The ratios between the areas of these rectangles should be close to the golden ratio [15] for a better appeal, i.e.,

$$\frac{Y_g}{Y_k} = \frac{Y_k}{Y_k + Y_g} = \phi, \quad Y_k > Y_g, \qquad (2)$$

where Y_k (red dotted line in Fig. 9(b)), Y_g (green dotted line in Fig.9(b)) denote the vertical extents of sky and support regions respectively and ϕ is the golden ratio. In order to maintain the aesthetic balance, these ratios should be equal to the golden ratio (ϕ), which is approximately equal to 1.61803. From the Fig. 8(a), these these ratios are observed to be 1.6011 and 1.5934 ($\approx \phi$), while in Fig. 8(b) the same numbers are 0.4533 and 0.6743, which makes the former image more appealing.

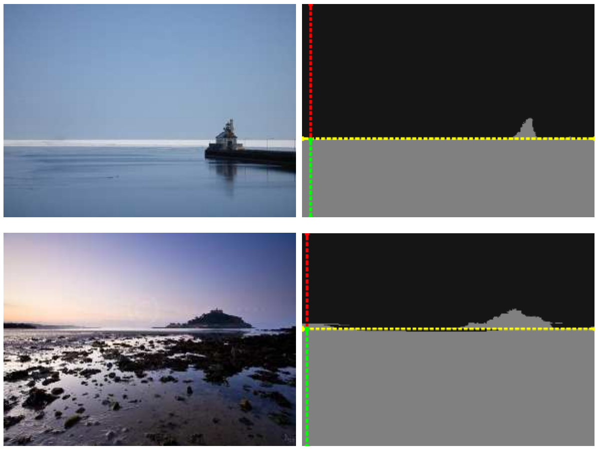


Figure 9: Visual weight measures for sky and support regions: column in the left shows original images, column in the right show the horizon (yellow dotted line) and the respective vertical extents (red and green dotted lines). Quantitative interpretation of these extents are provided in Table 2, with the first and second rows corresponding to the top-right and bottom-right images, respectively.

F_a	Visual Weight Deviations ($\mathbf{W}$)	
	$\lvert\phi - \frac{Y_g}{Y_k}\rvert$	$\lvert\phi - \frac{Y_k}{Y_k+Y_g}\rvert$
4.58	0.099	0.012
3.23	1.112	0.976

Table 2: Mapping user rated appeal factor (F_a) with the visual weight deviations from the Golden ratio ($\mathbf{W}$) for the two seascapes in Fig. 9

We make a reasonable assumption that the photographic frame is approximately aligned with the horizon so that Y_k and Y_g could be estimated by averaging the vertical extents of the pixels belonging to sky and support regions, respectively. The deviations of the individual ratios $\frac{Y_g}{Y_k}$ and $\frac{Y_k}{(Y_k+Y_g)}$ from the Golden Ratio (ϕ) form the aesthetic feature ($\mathbf{W}$) for photographs of seascapes or landscapes. Formally,

$$\mathbf{W} = \left[\left| \phi - \frac{Y_g}{Y_k} \right|, \left| \phi - \frac{Y_k}{Y_k + Y_g} \right| \right]. \tag{3}$$

The two high-level features discussed above are clearly not the only ones that can capture an image's aesthetic appeal. They were chosen because they can be reliably quantified using existing techniques and address typical photographs found in Internet photo collections. Other metrics from the photography literature are either too abstract, demanding a sophisticated understanding of the image scene that is beyond current computer vision algorithms, or would apply to only a relatively small subset of photographs.

2.4 Learning and Prediction

The aesthetic appeal for the two different types of photographic composition that we have addressed here can be associated with the features extracted using two smooth functions defined as:

$$f_{rf}(F_a) : R^4 \rightarrow R, R \in \mathbf{F} \tag{4}$$

$$f_{vw}(F_a) : R^2 \rightarrow R, R \in \mathbf{W} \tag{5}$$

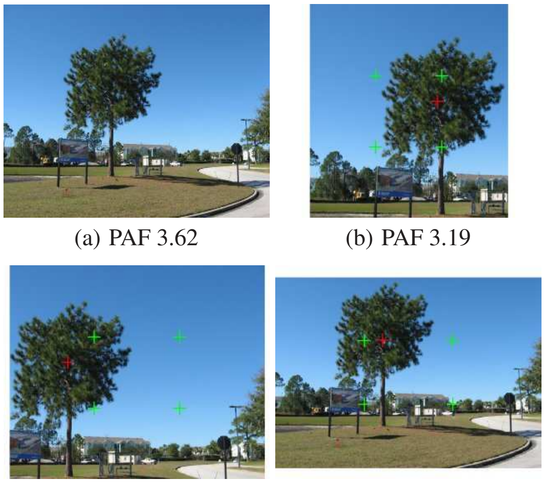
based on Eqns (1) and (3). We use two independent, soft-margin support vector regressors implemented using [10] to learn the above non-linear mappings. We employ a coarse grid search with the SVR's error parameter values (C) from $0.1, 1, 10$, and tube-width values (ϵ) from $0.01, 0.1, 1, 10$ on an RBF kernel with σ values from $0.5, 1, 2$. We select 150 random images from either composition class for training and use the rest for testing. The best prediction accuracy of $87.3 \pm 3\%$ for photographs with single foregrounds is reported for $\sigma = 2, C = 0.1, \epsilon = 1$. The same number for the latter composition category is reported to be $96.1 \pm 2\%$. A detailed quantitative analysis is provided in the results section.

3. ENHANCING COMPOSITION

Our recomposition technique is built upon inputs from the same aesthetic features that used to evaluate a given composition. We introduce two separate enhancement approaches for the two categories of compositions. The first aims to relocate the foreground object so as to increase the predicted appeal factor of the image while maintaining the scene integrity; i.e., an object on land remains in contact with the ground and does not float into the sky. The second focuses on increasing the appeal factor of landscape and seascape images by better balancing the visual weights of the sky and support regions. Both approaches are detailed in the following subsections.

3.1 Optimal object placement

The problem of spatial recomposing is closely related to the simpler task of optimally cropping a given photograph in order to enhance its visual appeal as studied by the authors in [18]. Since the locations of the stress points are determined entirely by the frame dimensions, one can crop a photograph to better align the dominant object with a given stress point, as shown in Fig. 10.



(a) PAF 3.62

(b) PAF 3.19

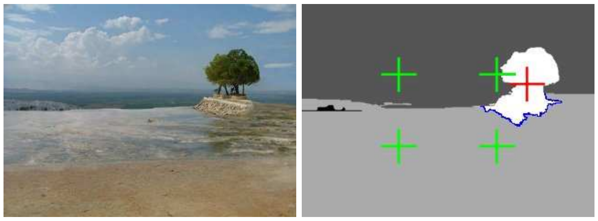
(c) PAF 3.48

(d) PAF 3.69

Figure 10: Illustrating an analogy between ideal positioning of the subject and optimally cropping the photograph: (a) original image; (b)–(c) cropped samples of the original image that move the visual attention center (centroid of the tree, denoted by a red cross-hair) towards/away from the stress points (green cross-hairs); (d) a near-optimal crop that aligns the visual attention center near the top-left stress point. For every crop, the respective appeal factor is determined using the relative foreground location feature based regressor.

Unfortunately, while this analogy prescribes a straightforward solution to the problem of optimal foreground alignment, it is unsatisfactory in two key respects. First, cropping reduces the size of the image frame and can alter its aspect ratio. Second, and more importantly, cropping can lead to the loss of valuable image information, such as key aesthetic features in the background. This motivates us to attempt a more ambitious goal: moving the foreground object in the image frame to a better location without compromising the semantics of the scene. In the context of Fig. 11, we seek to move the foreground object (tree) in such a manner that the predicted appeal factor after the relocation increases while keeping the tree in contact with its support in the background.

Recall $\mathbf{x_0}$ as the location of the current *visual attention center* (the foreground object's centroid in image coordinates), we define the support neighborhood for the foreground as ψ_w. In other words, these are the set of pixels that lie within $w \times w$ neighborhood of the boundary of the foreground. With a slight abuse of notation, let $\psi_w(\mathbf{x_0})$ denote the set of pixels forming the *support neighborhood* at the object's original location and $\psi_w(\mathbf{x})$ to be those pixels that would form the support neighborhood were the object mask to be centered at $\mathbf{x}$ rather than $\mathbf{x_0}$ at a single iteration. Clearly, the shape of the support neighborhood is constant for any $\mathbf{x}$, but the intensity values of the underlying pixels (in each of the three channels, assuming an RGB colorspace) from the background would change. Now, we express the problem of relocating the object to an aesthetically favorable location $\hat{\mathbf{x}}$ as the following optimization problem:



(a) (b)

Figure 11: Formulating the optimal object placement problem: (a) Original image; (b) dominant foreground object, sky and support regions are represented using white, dark gray, and light gray pixels respectively; Blue pixels are special cases of support pixels in the foreground object's neighborhood (ψ_w); four green cross-hairs mark the stress points; red cross-hair marks the visual attention center($\mathbf{x_0}$).

$$\arg\max_{\mathbf{x}} f_{rf}(F_a) \quad \text{s.t.}\lambda(\mathbf{x},\mathbf{x_0}) < \delta, \tag{6}$$

where δ is a human-specified real-valued number which enforces how closely the support regions must match, and $\lambda(\mathbf{x},\mathbf{x_0})$ is a smoothness term computed over the pixel intensities and gradients in the spatial neighborhoods of $\mathbf{x}$, $\mathbf{x_0}$ as:

$$\lambda(\mathbf{x},\mathbf{x_0}) = S_I + \beta S_\nabla. \tag{7}$$

Here β is a regularization parameter, usually set to a high value (∞) for regions with large texture variations, S_I and S_∇ are the intensity and gradient components of the smoothness term respectively, calculated as:

$$S_I = \sum_{\psi_w \forall\{R,G,B\}} ||I(\psi_w(\mathbf{x})) - I(\psi_w(\mathbf{x_0}))||_1 \, , \tag{8}$$

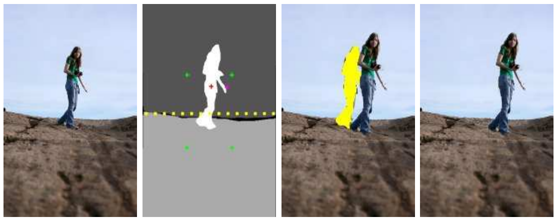
$$S_\nabla = \sum_{\psi_w \forall\{R,G,B\}} ||\nabla(\psi_w(\mathbf{x})) - \nabla(\psi_w(\mathbf{x_0}))||_1 \, . \tag{9}$$

The solution to Eqn. (6) gives us the new location for the visual attention center of the foreground object ($\hat{\mathbf{x}}$). We obtain $\hat{\mathbf{x}}$ by optimizing using standard techniques. Fig. 12 shows some intermediate outputs from our algorithm during the optimization process. We observe that the location of the horse shifts from frame to frame. In the best result, the location of the horse is well aligned with a stress point and the support neighborhood is highly consistent with that of the original image. We explicitly set the search window to a homogeneous grass-covered region, for a faster convergence.

We discuss whether (and how) to scale the horse to correct for perspective, and how to inpaint the hole left at its original location later in the paper. Given the small size of this optimization problem, we use an exhaustive search with a user-specified quantization size to optimize Eqn. (6) as this guarantees a globally optimal solution. Furthermore, we reduce the complexity of the search from O($h \times w$) to O($(h-l) \times (w-m)$) where l, m are dimensions of the region that is semantically least likely to contain the foreground after recomposition. A detailed qualitative and quantitative analysis of the recomposition technique is provided in Section 3.1.1.

3.1.1 Rescaling to Maintain Perspective

Simply translating the foreground object in the scene is insufficient for photorealistic recomposition. This is because moving an object vertically in the image changes the depth at which it is perceived in the scene. For instance, an object on the ground should shrink as it translates up in the image by a factor that depends on imaging characteristics such as the focal length and tilt of the camera. Thus, spatial recomposition must correctly rescale the object to maintain photorealism.



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

Figure 13: Spatial recomposition procedure: (a) Original image, (b) Corresponding segment map with light gray, dark gray, and white pixels denoting regions that belong to support, sky, and the object respectively; yellow dotted line showing the horizon; the four stress points are indicated by green cross-hairs while the location of the visual attention center is shown by the red cross-hair; the output of the optimal placement algorithm is the translated centroid of the object shown by a purple cross-hair. (c) The foreground object placed in the optimal location leaving a yellow hole in its original location. (d) Final result after inpainting.

Fortunately, we can employ methods that automatically estimate the location of the horizon in the image (e.g., [9]) to determine the correct size of the foreground object at its new location using the following straightforward equation:

$$v_x = \frac{D_x}{D_y}(v_y - y_2) + x_2, \tag{10}$$

where $\mathbf{v} = (v_x, v_y)$ is the vanishing point [8] i.e., the point of intersection between the horizontal line $y = v_y$ and the line through the original object location $\mathbf{x_0}$ and its modified location $\hat{\mathbf{x}}$. D_x/D_y is the slope of this line and x_2, y_2 are the components of $\hat{\mathbf{x}}$. The scaling factor is computed as:

$$f_s = \frac{||\mathbf{v},\mathbf{x_0}||_2}{||\mathbf{v},\hat{\mathbf{x}}||_2}. \tag{11}$$

For images where the vanishing line information cannot be reliably determined, we simply keep the size of the object constant (equivalent to orthographic projection). We show our results in Fig. 13, where Fig. 13(d) shows a slight increase in size as the foreground object moves towards the viewer in the image frame. A fast bicubic interpolation algorithm is applied to perform the scaling operation. More results are shown in Fig. 18.

3.2 Balancing visual weights

For scenes that have a clearly demarcated horizon or vanishing line (refer Fig. 14), we can apply spatial recomposition to better balance the visual weight of the sky to the frame. Let us assume that a horizontal line divides our image in the ratio $\frac{Y_k}{Y_g}$. A fixed-step Y_k expansion or contraction strategy can be applied here to solve Eqn. (5) which leads to the optimal combination of visual weights that maximizes the appeal factor.

Since this is relatively less complex than solving for the optimal foreground placement location, we resort to a simpler technique by assuming the following holds good at the optimal solution:

$$\frac{Y_k}{Y_g} = k\frac{Y_g}{Y_k + Y_g}, \quad k > 0. \tag{12}$$

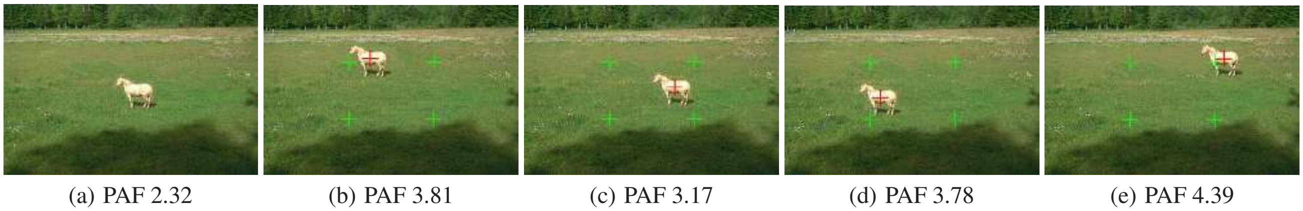
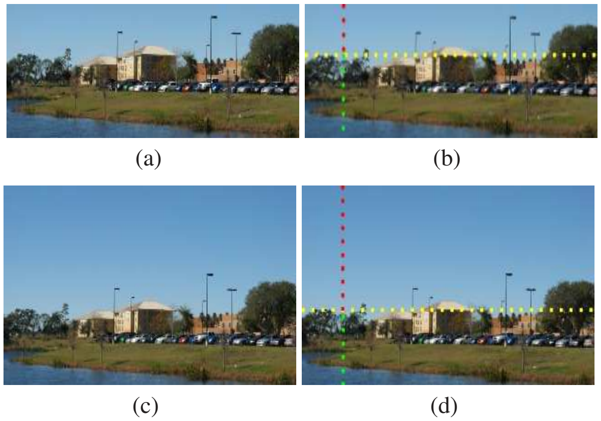

(a) PAF 2.32 (b) PAF 3.81 (c) PAF 3.17 (d) PAF 3.78 (e) PAF 4.39

Figure 12: Intermediate results from spatial recomposition with corresponding predicted appeal factor computed by the rule of thirds based regressor: (a) original image; (b) - (d) potential solutions for relocating foreground object; (e) optimal location. Note that these results do not include rescaling the object in accordance with perspective as described in Section 3.1.1.


(a) (b)

(c) (d)

Figure 14: Altering a composition to balance visual weights: Which image in the left column looks more appealing? (a) Original image. (b) Corresponding image showing the distribution of visual weights. (c) The vertical extent of sky increased using our method to balance the distribution of visual weights, improving the overall aesthetic appeal of the image; (d) Modified distribution of the visual weights.

Let h be the vertical extent that Y_k must be increased so that:

$$\frac{Y_k + h}{Y_g} = \frac{Y_g}{(Y_k + h) + Y_g}. \tag{13}$$

With a couple of algebric substitutions in Eqn. (13) from Eqn. 12, we obtain a quadratic equation in h, which can be easily solved for two values of h. A positive value of h indicates an increase of Y_k by h, while a negative value of h means decrease of Y_g by h, leading to an increase or decrease in the overall image height. In order to increase the height of the image, we are required to in-paint the newly-added region with information available from neighboring pixels. Decreasing the height is simply performed by cropping the image appropriately. For inpainting, we employ the straightforward patch-based region filling algorithm proposed by Zhang *et al.* [22]. We limit the search for target patches in 20×20 neighborhood of the source patch. For most of the images in our dataset, we achieve aesthetically pleasing results with fewer than 60 iterations of a graphcut-based patch updating mechanism discussed in [22]. However, the algorithm frequently introduces minor artifacts into the background of our recomposed images, that require interactive retouching.

4. EXPERIMENTAL RESULTS

We performed an extensive qualitative and quantitative evaluation of the proposed methods, summarized as follows. We apply the proposed recomposition techniques separately to 200 images taken from both categories(single object compositions and sea/land-scapes) of the dataset. This is facilitated by a graphical tool where a user is interactively asked to label regions sky, support or the foreground object using closed polygons. An automatic segmentation option is also provided which can be used for relatively less complex scenes, for example scenes without shadows, reflection etc. Once the user is satisfied with the segmentation process, he/she chooses which algorithm to apply. Depending on the algorithm selected, the tool employs either of the two techniques discussed in Subsection 3.1 or Subsection 3.2.

Of the 200 images in the single subject composition category, 38 have appeal factors in the interval $(1, 2]$, 49 in $(2, 3]$, 75 in $(3, 4]$, and the rest are in the last interval $(4, 5]$. The recomposited images are then evaluated by users in the same way as discussed in Subsection 2.2. We observe a clear increase in aesthetic appeal of images whose F_a values were in the $(1, 2]$ and $(2, 3]$ intervals as sectors corresponding to these intervals shrink in the rightward pie-chart in Fig. 15. The increased area of the sector corresponding to the interval $(3, 4]$ in the same pie-chart show in favor of the argument that some images from the lower intervals have moved up, after recompositing. Since the aggregated statistics shown in the pie-chart do not provide insight on how individual images could have been affected as a result of the process, we also plot the average appeal factors of images in each interval, before and after recompositing.

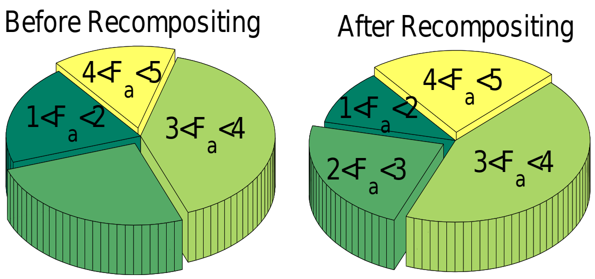


Figure 15: Qualitative results on images recomposited using the optimal object placement(refer to subsection 3.1) technique. In both pie-charts, each sector represents the fraction of images whose respective appeal factor lie in one of the four discrete intervals($(1, 2]$, $(2, 3]$, $(3, 4]$, $(4, 5]$). Recompositing shows definite improvement in the lower two intervals as their respective sectors shrink in the right pie-chart. The bar-chart in the bottom shows the net improvement of appeal factors pertaining to each intervals after recompositing.

A similar experiment is performed for the land/sea scape images. In this case, we begin with 82 images whose appeal factors are in the interval $(1, 2]$, 86 in $(2, 3]$, 21 in $(3, 4]$, and the rest in $(4, 5]$. We see a similar trend as observed in Fig. 15 in this setting (Fig. 16) as well. The bars corresponding to the interval $(4, 5]$ indicate that there is little scope for improvement for images that are already aesthetically appealing.

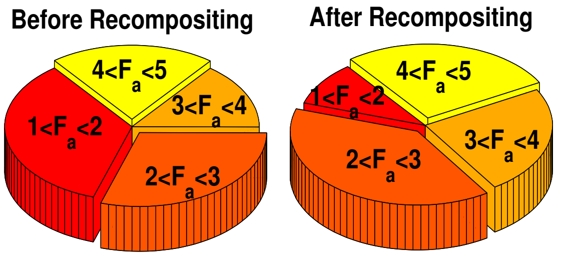


Figure 16: Qualitative results on images recomposited using visual weight balancing(refer to subsection 3.2) technique. We observe a similar trend as seen in Fig. 15 in the increase of appeal factors for images recomposited using the visual weight balancing technique. Refer to the text for details.

Some qualitative results obtained after recomposition are given in Fig. 18. Note how the scales are adjusted for foregrounds in some of the images (person, cow, building, boat) with inputs from user about the respective scenes. The bottom two rows show some results after applying the visual weights based recomposition. Unlike Figs 1(c) and 1(d), where the non-sky region is cropped optimally to increase the visual appeal, these images show results of sky-region augmentation to increase the appeal. Fig. 17 shows some cases where the proposed method either reduces the visual appeal or makes negligible improvements.

5. CONCLUSION

We have introduced a new multimedia application that enables users to assess the aesthetic quality of a photograph using geometric rules of composition, and then to make an informed decision on how to improve the photograph using spatial recomposition. Rather than prescribing a fully-automated solution, we allow user-guided object segmentation and inpainting to ensure that the final photograph matches the user's criteria. Our approach achieves 86% accuracy in predicting the attractiveness of unrated images, when compared to their respective human rankings. Additionally, 73% of the images recomposited using our tool are ranked more attractive than their original counterparts by human raters.

In future work, we plan to replace the resizing operations currently used in recomposition by a more context-aware resizing algorithm [1]. Although this paper demonstrates results only on two common classes of photo compositions (single subject and landscape/seascape), our ideas extended naturally to compositions involving multiple foreground objects. A segmentation algorithm with minimal intervention could generate additional training data for a robust aesthetic model that could be applied to improving Internet image search. In addition, our enhancement technique could be applied synergistically with low-level image editing techniques [14], while preserving the semantic essence of the scene.

6. REFERENCES

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. In *SIGGRAPH*, 2007.

[2] S. Banerjee and B. Evans. Unsupervised automation of photographic composition rules. In *SPIE Sensors, Color, Cameras, and Systems for Digital Photography*, 2004.

[3] M. Boutell and J. Luo. Bayesian fusion of camera metadata cues in semantic scene classification. In *CVPR*, 2004.

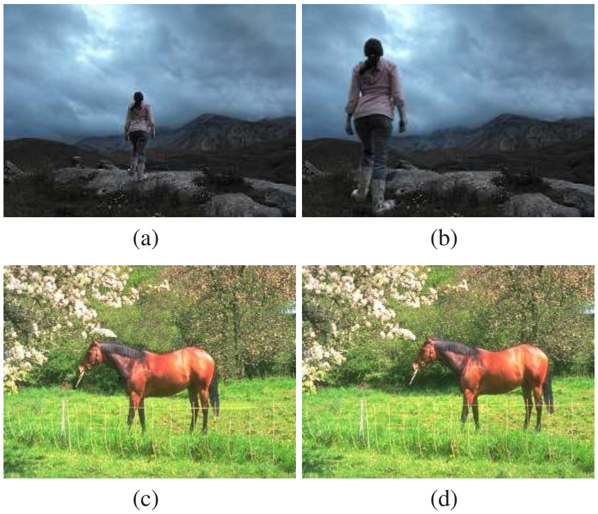


Figure 17: Failure cases of our spatial recomposition technique: (a) Original image looks better than the (b) Recomposed version. (c) and (d) look visually very similar to each other.

[4] T. S. Cho, M. Butman, S. Avidan, and W. T. Freeman. The patch transform and its applications to image editing. In *CVPR*, 2008.

[5] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik. Image quality assessment based on a degradation model. *IEEE Trans. Image Proc.*, 9(4), 2000.

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.

[7] R. Datta, J. Li, and J. Z. Wang. Learning the consensus on visual quality for next-generation image management. In *ACM MM*, 2007.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[9] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75, 2007.

[10] T. Joachims. Making large-scale SVM learning practical. In *Advances in kernel methods: support vector learning*, 1999.

[11] P. Jonas. *Photographic composition simplified*. Amphoto Publishers, 1976.

[12] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, 2006.

[13] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. In *SIGGRAPH*, 2008.

[14] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Trans. PAMI*, 30(2), 2008.

[15] M. Livio. The golden ratio and aesthetics. *Plus Magazine — Living Mathematics*, 22, Nov. 2002.

[16] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV*, 2008.

[17] A. Mansoor, M. Haider, A. Mian, and S. Khan. A hybrid image quality measure for automatic image quality assessment. In *Scandinavian Conf. Image Analysis*, 2009.

[18] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *ACM MM*, 2009.

[19] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *ACM MM*, 2009.

[20] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19, 2006.

[21] J. You, A. Perkis, M. Hannuksela, and M. Gabbouj. Perceptual quality assessment based on visual attention analysis. In *ACM MM*, 2009.

[22] Y. Zhang, J. Xiao, and M. Shah. Region completion in single image. In *Proc. EUROGRAPHICS*, 2004.

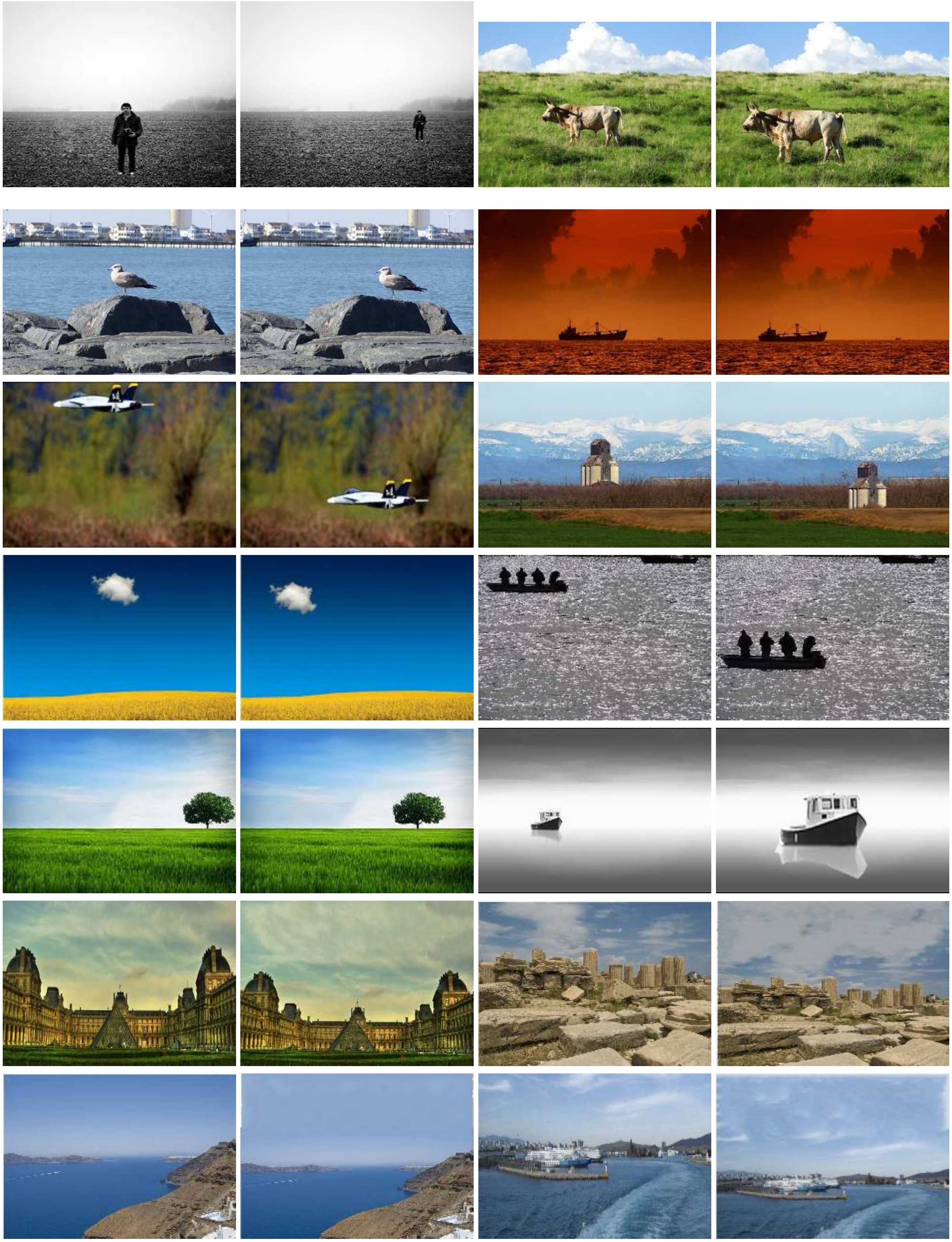


Figure 18: Results of spatial recomposition on a subset of images from our dataset (Success): Each pair of images has the original image on the left and its recomposed counterpart on the right. In the top five rows, we have images recomposited using the optimal object placement algorithm, whereas for the bottom two rows, visual weights are optimally altered for a visually appealing effect.