Enhancing Visual Dominance by Semantics-Preserving Image Recomposition

Lai-Kuan Wong
Department of Computer Science
National University of Singapore
Singapore 117417
lkwong@comp.nus.edu.sg

Kok-Lim Low
Department of Computer Science
National University of Singapore
Singapore 117417
lowkl@comp.nus.edu.sg

ABSTRACT
We present a semi-automatic photographic recomposition approach that employs a semantics-preserving warp of the input image to enhance the visual dominance of the main subjects. Our method uses the tearable image warping method to shift the subjects against the background (and vice versa), so that their visual dominance is improved, and yet preserve the desired spatial semantics between the subjects and the background. The recomposition is guided by a measure of the resulting visual dominance of the main subjects. Our user experiment shows the effectiveness of the approach.

Categories and Subject Descriptors
I.4.9 [Image Processing and Computer Vision]: Applications

General Terms
Algorithms, Experimentation, Human Factors

Keywords
Computational aesthetics, photographic recomposition, tearable image warping, image semantics, visual dominance

1. INTRODUCTION
In photography, it is essential to make the subjects of interest dominant so that the viewers' attention is directed to what the photographer wants them to see. To increase the subjects' visual dominance (which usually also increases photographic aesthetics), professional photographers apply a set of well-known techniques to make a subject stands out from the background, such as by selective focusing using shallow depth of field, by choosing a viewpoint and framing to avoid cluttered background and merger, and by enhancing the lighting on the subject. In Figure 1, we can observe that the giraffe in the image captured by a professional photographer is distinctively more visually dominant due to a good choice of camera viewpoint.

Recent state-of-the-art automatic recomposition methods have achieved some success in improving aesthetics of images [2, 4, 5, 6]. However, almost all these methods work well only when the subjects are already visually dominant with respect to their immediate background. None has attempted to make the subjects more dominant by directly changing the subject-background spatial relationship.

MM'12, October 29–November 2, 2012, Nara, Japan.

Figure 1: Photographs captured by (left) a professional photographer, and (right) a casual photographer.

We propose a semi-automatic recomposition approach to enhance the visual dominance of main subjects. We employ tearable image warping [8], which can change subject-background spatial relationship while preserving scene consistency, to transform the input image. To guide tearable warping to increase the dominance of the user-selected subjects, we formulate a visual dominance energy based on a simplified version of the Itti-Koch visual attention model [3]. The energy function is used to maximize the contrast of two low-level features—intensity and color—between a subject and its immediate background. Figure 2 shows a result from our recomposition method and its corresponding saliency map [3]. Our method has effectively transformed the input image to increase the image saliency of the subjects, thereby increased their visual dominance.

2. RELATED WORK
Automatic aesthetics image recomposition is a relatively new field, with a handful of prominent research work. Nishiyama et al. [6] proposed the first aesthetics-based cropping technique. A trained aesthetics assessment model was used to maximize aesthetics of the resulting cropped image. Based on well-grounded photo composition rules, Liu et al. [4] proposed a computational means to measure composition aesthetics and utilized a crop-retarget operator that combines pure warping and cropping to recompose an image. Following this, Liu et al. [4] later proposed a real-time warping-based aesthetics retargeting tool by reformulating the aesthetics score computation. Bhattacharya et al. [2] developed a framework for photo-quality assessment and enhancement based on visual aesthetics. They trained their aesthetics scoring system using real user data and provided a framework to support semi-automatic patch relocation. All these methods measure aesthetics using common rules such as rule-of-thirds, visual balance and diagonal dominance. Little or no emphasis is put into enhancing the dominance of photo subjects.

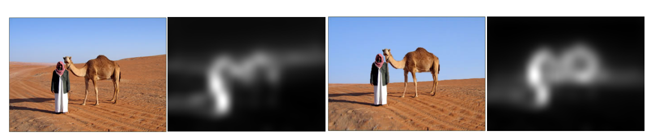

Figure 2: (Left) Input image and its saliency map. (Right) Result image from our method and its saliency map.

3. MAKING THE SUBJECT DOMINANT

The core idea of our recomposition approach is to geometrically transform the immediate background of a photo subject, such that the subject becomes dominant. Often, the resulting effect is analogous to a change of the camera viewpoint. To achieve this objective, we require two important components: (1) an *image operator* to transform the input image, and (2) a *subject dominance measure,* to guide the image operator to produce an output image with enhanced subject dominance.

We employ *tearable image warping* [8] as the image operator, largely due to its ability to preserve image semantics while allowing significant change in image composition. Tearable warping divides an object boundary into *tearable* and *non-tearable* segments. The non-tearable segments normally correspond to the real contacts in 3D world, and the original background must remain attached to the segments. These segments are represented by *object handles*, which are polylines drawn by the user to specify the non-tearable segments. On the other hand, tearable segments allow objects to break away from its background so that the background can be warped more independently while keeping the objects undistorted.

We formulate the *subject dominance measure* based on a simplified adaptation of the Itti-Koch biological-inspired visual attention model [3]. We treat the photo subject as the center and its immediate background as the surround, and measure subject dominance by computing the center-surround differences of two low-level features—intensity and color—between the subject and its background. For this, the Lab color space is ideal since the L*, a* and b* components represent the luminance, the color position between red and green (R-G) and the color position between yellow and blue (Y-B) respectively. A highly dominant subject should exhibit high center-surround contrast in one or more of these intensity and color features.

For efficiency purposes, in our implementation, we adopt the triangle-based warping method where a triangle mesh is used to represent the background image. Ideally, center-surround differences for the L*, a* and b* features should be measured between neighboring triangles connecting the subject and its immediate background. However, as the subject is not warped with the background and the set of background triangles neighboring the subject can change dynamically during optimization, it is not efficient to represent the object as a triangle mesh and recompute its new neighboring background triangles for each iteration of the optimization. Therefore, we compute the center features by finding the average L*, a* and b* features for the whole subject. The set of surround features consists of the L*, a* and b* features for each triangle in the background within an expanded bounding box of the object. Using this center-surround contrast measure to guide tearable image warping, our algorithm makes the photo subject more dominant by enlarging background triangles that exhibit high contrast with the photo subject while compressing background triangles with low center-surround contrast. The synthetic examples in Figure 3 illustrate the effectiveness of our approach to increase the contrast of the rectangular subject. Results of our approach on natural images (e.g. Figure 4) have shown this approximation to be still effective.

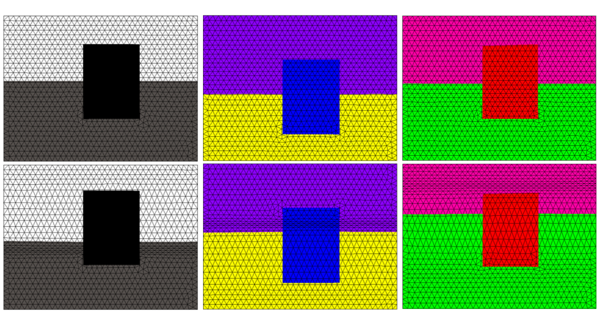

Figure 3: (Top) Synthetic input images that possess luminance and color contrast. (Bottom) Results of our algorithm show increased visual dominance of the rectangle.

4. ALGORITHM

As illustrated in Figure 5, our algorithm is performed in three main steps: (1) *decomposition*, (2) *warping and subject dominance optimization*, and (3) *image compositing*. In the decomposition step, an image is separated into a background layer and one or more object layers. Holes left by the objects are automatically inpainted [1]. Next, we perform warping on the inpainted background image to maximize the dominance of the photo subjects. The algorithm is detailed in the following sub-sections. In the final image compositing step, objects are pasted back to the warped background at the new positions of their handles.

4.1 Subject Dominance Energy

To maximize the luminance and color contrast in the Lab color space, we minimize the *subject dominance energy*, which is defined as

$$E_D = E_L + E_C, \tag{1}$$

where E_L is the *luminance contrast energy*:

$$E_L = \sum_{o \in O} \sum_{t \in T_O} s_t (|L'_t - L_o| - \psi_L), \tag{2}$$

and E_C is the *color contrast energy*:

$$E_C = \sum_{o \in O} \sum_{t \in T_O} s_t \left(\sqrt{(a'_t - a_o)^2 + (a'_t - b_o)^2} - \psi_{ab}\right). \tag{3}$$

T_O is the set of background triangles of the object. L'_t, a'_t and b'_t are the average values of L*, a* and b* of each triangle $t \in T_O$ in the output mesh, and similarly, L_o, a_o and b_o are the average values for each object $o \in O$. s_t is the area scaling to be applied to each original triangle $t \in T_O$. It is set to $s_t = s_t^y$ if scaling of triangles is allowed only in the vertical dimension, or $s_t = s_t^x s_t^y$ if in both vertical and horizontal dimensions. We use only vertical scaling in our implementation. ψ_L and ψ_{ab} are parameters to indicate the target level of dominance, where

$$\psi_L = \left(\frac{1}{|T_O|} \sum_{t \in T_O} |L_t - L_o|\right) + \mu, \tag{4}$$

and

$$\psi_{ab} = \left(\frac{1}{|T_O|} \sum_{t \in T_O} \sqrt{(a_t - a_o)^2 + (b_t - b_o)^2}\right) + \mu, \tag{5}$$

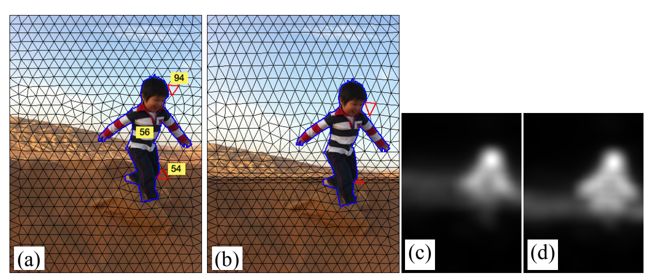

Figure 4: (a) Input image with triangle mesh. Yellow rectangles show the average luminance (L*) values in the subject and the red triangles. (b) Output image. Triangles have been expanded and compressed accordingly. (c & d) Saliency maps from input and output images respectively.

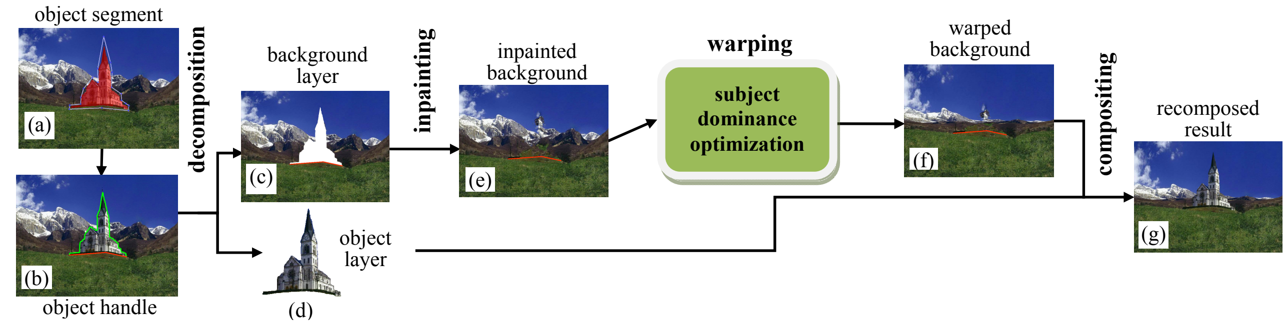

Figure 5: Steps in our method. In (b), the object boundary is shown in green and the object handle in red.

where L_t, a_t and b_t are the average values of L*, a* and b* of each triangle $t \in T_O$ in the input mesh. μ is a parameter related to the target level of dominance and its value varies for each image, depending on two features—background contrast and subject-background contrast. To obtain a μ value, we create a set of synthetic images with fixed background contrast but varying subject-background contrast. For each image, we then find the μ value that produces result with the maximum subject-background difference, as illustrated in Figure 6. By plotting the maximum μ value against the standard deviation of subject-background difference and fitting a graph to this plot, we obtain a formula for μ. We repeat this process for a few sets of images with each image set having a different background contrast. We then find a graph that fits reasonably well to each plot and obtain the following adaptive μ value:

$$\mu = \frac{S(F)B(F)}{\tau} - \frac{B(F)^2}{18*S(F)}, \quad (6)$$

where F is the feature (L* or ab*), $S(F)$ is the subject-background contrast, given by the standard deviation of the center-surround feature differences, and $B(F)$ is the background contrast, given by the standard deviation of the background triangle features. The value of τ depends on the features: for luminance, $\tau = 9$ and for color feature, $\tau = 6$.

4.2 Warping Energy

The warping energy consists of a *scale transformation error*, E_w and a *smoothness error*, E_s [6, 8]. The scale transformation error favors non-uniform scaling (as opposed to general affine transformation) while the smoothness error tries to avoid discontinuity by minimizing the scale difference between neighboring triangles. The *handle shape constraint* rigidly preserves the shape and orientation of the object handles. The *image boundary* and *object boundary constraints* ensure that the original image boundary remains on the boundary and user-defined objects do not move out of the image boundary. We also limit the scale factor of the background triangles to a minimum value of 0.15 to avoid triangle fold-over and over-compression.

4.3 Total Energy

To enable automatic recomposition, the subject dominance energy is combined with the warping energy:

$$E = \alpha E_w + \beta E_s + \gamma E_D \quad (7)$$

where E_w is the scale transformation error, E_s the smoothness error, and E_D the subject dominance energy. Minimizing the total energy is achieved by warping the input triangle mesh.

5. IMPLEMENTATION

We use a semi-automatic approach to obtain object segments from users. With the assistance of GrabCut [7], users can easily segment the objects by drawing a polygon around the object with a few clicks. Similarly, users can specify a polyline to define an object handle. We use the CVX Matlab toolbox to find the solution to the convex quadratic function defined in Eq. (7). Weights for the total energy, α, β and γ, are set to 1, 0.5 and 1, respectively. The handle shape and image boundary constraints are set as hard constraints while the object boundary and fold-over avoidance constraints are set as inequality constraints.

6. RESULTS AND USER STUDY

Our results in this paper were generated on a laptop with Intel Core2 Duo CPU 2.53GHz. Excluding time for inpainting, it took about 10 seconds to produce a recomposed image with resolution 800 x 600. Some results from our algorithm are shown in Figures 2, 4 and 7. We observe that many of the resulting images appear to be a change of viewpoint from that of the input images.

For an objective evaluation, we conducted an online user experiment. We let each participant compare 30 pairs of images. The images were collected from Flickr.com. Each time, an input image and our result were shown side-by-side, with left-right positions randomly chosen. The participants were instructed to choose one where the photo subject stands out more against the background.

The experiment had 40 participants, consisting of males and females aged between 22 and 46. The outcome is, for 83% of the image pairs, our results were chosen.

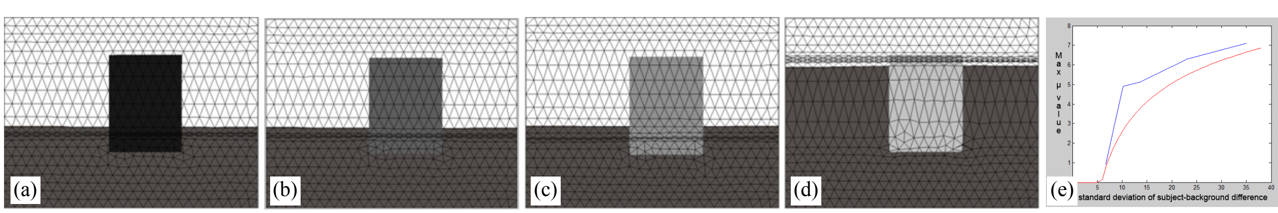

Figure 6: (a–d) Max μ value for each image with the same background contrast but varying subject-background contrast is found by compressing each image to the maximum. (e) (Blue) The plot of max μ value (y-axis) against the standard deviation (x-axis) of subject-background contrast. (Red) We obtain a formula for μ value by fitting a logarithm function to the plot.

7. MORE AESTHETICS FEATURES

Our recomposition approach can be extended to incorporate other aesthetics features to maximize image aesthetics.

Rule-of-Thirds The *rule-of-thirds energy* comprises the *power-point energy* that pulls subjects toward one of the four power points (the intersections of the two vertical and two horizontal power lines that divide the image into nine equal rectangular regions) and the *horizon energy* that pulls the horizon towards one of the two horizontal power lines. The rule-of-thirds energy, E_G is defined as

$$E_G = \sum_{o \in O} A_t D_P(o) \tag{8}$$

where $D_P(o)$ is the minimum distance from the subject centroid to the four power points and A_t is the normalized subject size.

The *horizon energy*, E_H tries to minimize the distance between the new horizon, $\hat{H}$ and the two power lines, PH_j, and is defined as

$$E_H = A_t \left(\min_{j=\{1,2\}} \left(\left| \hat{H} - PH_j \right| \right) \right). \tag{9}$$

If horizon does not exist, then $E_H = 0$.

Visual Balance The *visual balance energy*, E_v is computed as

$$E_v = \sum_{o \in O} A_t \left\| \hat{C}(I') - \hat{C}(o_i) \right\|_1 \tag{10}$$

where $\hat{C}(I')$ is the image center and $\hat{C}(o_i)$ is the weighted centroid of object *i*. For images with only one object, $E_v = 0$.

Total Aesthetics Energy For aesthetics optimization, we replace the subject dominance energy, E_D in Eq. (1) with the total aesthetics energy, E_A, which is defined as

$$E_A = w_D E_D + w_G E_G + w_H E_H + w_v E_v \tag{11}$$

where w_D, w_G, w_H and w_v are the weights for each aesthetics energy. In our implementation, they are set as w_D=1, w_G=2, w_H=0.1 and w_v=0.5.

Results Figure 8 shows some recomposition results produced using a combination of subject dominance and the above aesthetics features. Our aesthetics energy functions are able to guide the recomposition to adhere to the specific photographic rules.

8. CONCLUSION

We have demonstrated a successful automatic recomposition method that employs tearable image warping as the image operator and uses a simplified center-surround contrast measure to guide the warping to enhance the visual dominance of the photo subjects in the recomposed images. Our results and user experiment have shown the effectiveness. Moreover, our method can be extended with more aesthetics features.

9. ACKNOWLEDGEMENTS

This work is supported by Singapore MOE Academic Research Funds (Project Number: T1 251RES0804 and T1 251RES1104).

10. REFERENCES

[1] A. Criminisi, P. Pérez, and K. Toyama. Object Removal by Exemplar-Based Inpainting. IEEE Computer Vision and Pattern Recognition, pp. 721–728, 2003.

[2] S. Bhattacharya, R. Sukthankar, and M. Shah. A Framework for Photo-Quality Assessment and Enhancement Based on Visual Aesthetics. ACM Multimedia, pp. 271–280, 2010.

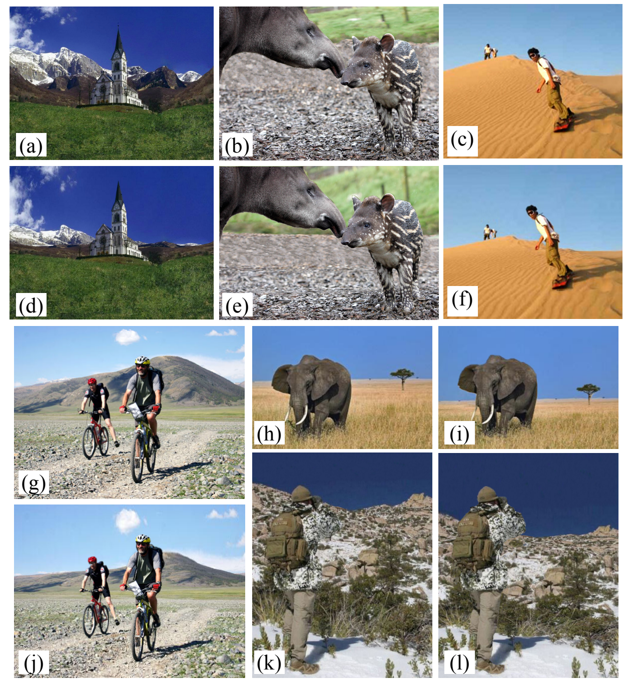

Figure 7: (a–c, g–h, k) Input images. (d–f, i–j, l) Results from our algorithm with only subject dominance energy.

[3] L. Itti, C. Koch, and E. Niebur. A Model of Saliency Based Visual Attention for Rapid Scene Analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(11): 1254–1259, 1998.

[4] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or. Optimizing Photo Composition. Computer Graphics Forum (Proceedings of Eurographics), 29(2): 469–478, 2010.

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

[6] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-Based Photo Cropping. ACM Multimedia, pp. 669–672, 2009.

[7] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2004), 23(3): 309–314, 2004.

[8] L.-K. Wong, and K.-L. Low. Tearable Image Warping for Extreme Image Retargeting. Computer Graphics International, 2012.

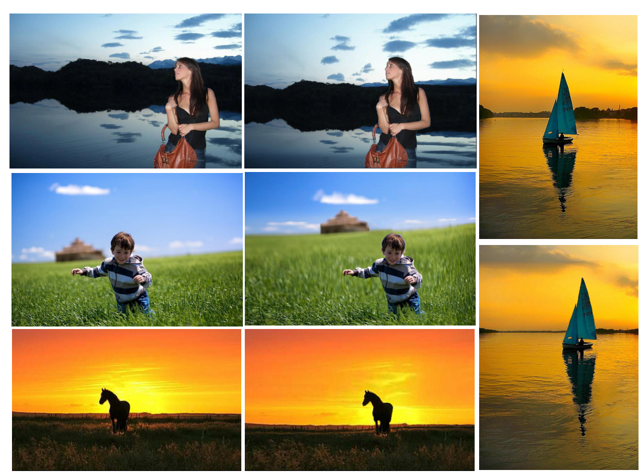

Figure 8: (Left column & top-right) Input images. (Middle column & bottom-right) Results from our algorithm using all aesthetics features.