R2P: Recomposition and Retargeting of Photographic Images

Hui-Tang Chang
Po-Cheng Pan
EE, National Taiwan University, Taipei, Taiwan
r02921089@ntu.edu.tw
b00901087@ntu.edu.tw

Yu-Chiang Frank Wang
Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan
ycwang@citi.sinica.edu.tw

Ming-Syan Chen
Electrical Engineering, National Taiwan University, Taipei, Taiwan
mschen@cc.ee.ntu.edu.tw

ABSTRACT

In this paper, we propose a novel approach for performing joint recomposition and retargeting of photographic images (R2P). Given a reference image of interest, our method is able to automatically alter the composition of the input source image accordingly, while the recomposed output will be jointly retargeted to fit the reference. This is achieved by recomposing the visual components of the source image via graph matching, followed by solving a constrained mesh-warping based optimization problem for retargeting. As a result, the recomposed output image would fit the reference while suppressing possible distortion. Our experiments confirm that our proposed R2P method is able to achieve visually satisfactory results, without the need to use pre-collected labeled data or predetermined aesthetics rules.

Categories and Subject Descriptors

I.4.3 [**Image Processing & computer vision**]: Enhancement; I.4.9 [**Image Processing & computer vision**]: Applications; I.5.4 [**Pattern Recognition**]: Applications - Computer Vision

General Terms

Algorithms, Experimentation

Keywords

Photography Composition; Graph Matching; Image Aesthetics; Computational Photography

1. INTRODUCTION

Image enhancement aims at improving the visual quality of an image, which can be achieved by applying photographic techniques (e.g., high dynamic range (HDR) or auto white balance (AWB)) or image processing algorithms (e.g., denoising or deblurring). However the only thing which cannot be easily altered is the geometric structure of visual components, i.e., *composition*, of an image.

Figure 1: Example of a source image which is recomposed and retargeted to fit the reference image.

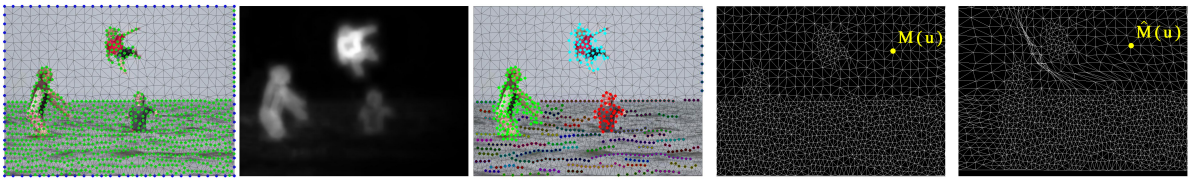
Improving image aesthetics is an alternative way to increase user satisfaction. Several computer vision and machine learning techniques have been applied to address this task. Since composition of an image has been viewed as a factor in assessing the image aesthetics, some researchers addressed the task of photographic composition [10, 9, 1, 6, 7]. Most of them aim at cropping the input image for satisfying particular rules of image aesthetics. For example, Nishiyama et al. [7] proposed sensation-based cropping for learning image aesthetics models. Santella et al. [8] utilized eye-tracking data, while Liu et al. [6] used pre-defined photography composition rules for image cropping. In addition to image cropping, Wong and Wong [9] proposed an algorithm to adjust the ratio of foreground/background regions. To perform recomposition, Bhattacharya et al. [1] trained their aesthetics models using pre-collected labeled (scored) image data, while Zhang et al. [10] further took the relationship between visual components into consideration.

Nevertheless, the above methods require pre-determined rules or pre-collected training data for performing photography recomposition. In practice, one cannot apply such rules to images taken in different scenarios. Thus, the robustness of such methods would not be sufficient. Recently, Chang et al. [2] proposed a transfer-learning based method for recompose a source image based on a reference image of interest. However, their approach applied image inpainting techniques which are computationally expensive. Moreover, they need to fine-tune their output image by post-processing the recomposed foreground regions.

Inspired by a mesh-based image retargetting work [5], we extend the work of [2] by integrating image retargeting and warping into a unified framework for image recomposition. As illustrated in Figure 1, we not only transfer the photo-

MM'15, October 26-30, 2015, Brisbane, Australia

DOI: http://dx.doi.org/10.1145/2733373.2806366.



(a) Mesh Generation (b) Salience Map (c) Object Extraction (d) Initial Mesh (e) Recomposed Mesh

Figure 2: Illustration of foreground extraction and recomposition. Note that $M(u)$ and $\hat{M}(u)$ denote the source and recomposed locations of the uth mesh point in the background, respectively.

graphic composition from the reference to the source image, the output image will be simultaneously retargeted to fit the reference. Thus, we refer to our proposed methods as recomposition and retargeting of photographic images (R2P).

In Section 2, we will present the details of our proposed method for joint recomposition and retargeting. We will discuss that, while no user interaction or pre-collected labeled training data are required, why our method is able to preserve image context information after the R2P process without additional post-processing procedures. It is worth repeating that, our work focus on transferring the photography composition from a reference image to the source image, rather than improving image aesthetic scores for the input image by pre-determined rules or features.

The contributions of our work are highlighted below:

- Automatically transfer the photography composition from a reference to the source image by solving a graph-matching based optimization problem.
- With the observed composition information, we perform a joint recomposition and retargeting on the source image, so that it would visually fit the reference image.
- Our proposed method is fully automatic, without the need to use pre-collected labeled training data, predetermine aesthetics rules, or user interaction.

2. OUR PROPOSED FRAMEWORK

2.1 Foreground Object Retrieval

Given a source and reference image pair, we first identify the foreground objects in both images. Prior works like [2] and [10] applied saliency detection and graph-cut for iteratively detect the foreground regions. For simplicity, we consider an one-pass algorithm as described below.

Inspired by a mesh-based image retargeting algorithm [5], we apply Canny edge detectors to detect the edges in an image, followed by sampling the pixels along the detected edges as the feature points. We note that, as depicted in Figure 2a, we mark the features points of detected edges and boundaries in green and blue, respectively. In order to describe smooth regions (e.g., sky in Figure 2a), we additionally add random feature points for collecting a sufficient number for representation purposes. Once such feature points are obtained, we utilize Delaunay triangulation algorithm to construct triangular meshes, as shown in Figure 2a.

Instead of iteratively applying saliency detection for determining possible foreground objects, our method directly combines the derived meshes and the saliency detection output of [4] (e.g., Figure 2b) for separating foreground and background regions, which is achieved by DB-scan clustering. Finally, N foreground objects set can be automatically observed. Also, we record all strong edges detected in the background regions (see Figure 2c), which will be crucial for the final stage of joint recomposition and retargeting.

2.2 Recomposition via Graph Matching

In our work, we apply a *composition graph* $G = (V, E, A)$ to describe each foreground object, and the recomposition can be performed by graph matching. For each graph, V and E denote the foreground objects and the edges connecting between them, respectively. We have A contain the attributes a_s for each vertex and $\mathbf{a}_l$ for each edge. That is, for $V_x \in V$, $a_s(x)$ indicates the area of that region. For each edge $e_{xy} \in E$, we have $\mathbf{a}_l(x, y)$ measuring the normalized Euclidean distance between the two vertices of edge e_{xy}.

To perform photography recomposition, we solve a graph-matching based optimization problem based on the observed affinity matrix $\mathbf{W}$ between the two composition graphs G^S and G^R, which are the constructed by source and reference images, respectively. The affinity matrix $\mathbf{W}$ is composed of the affinity between vertices and edges in the dimension of $|V^S||V^R| \times |V^S||V^R|$. In $\mathbf{W}$, each entry w is defined as:

$$w_{xm;yn} = \begin{cases} f(v_x^S, v_m^R) & \text{if } x = y,\ m = n \\ f(e_{xy}^S, e_{mn}^R) & \text{if } x \neq y,\ m \neq n \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where the affinity functions are calculated as follows:

$$\begin{aligned} f(v_x^S, v_m^R) &= min[a_s(x), a_s(m)]/max[a_s(x), a_s(m)], \\ f(e_{xy}^S, e_{mn}^R) &= 1/\|\mathbf{a}_l^I(x, y) - \mathbf{a}_l^R(m, n)\|. \end{aligned} \quad (2)$$

Now, we convert the original recomposition problem into the task of determining an indicator matrix $\mathbf{Z} \in \{0, 1\}^{|V^S| \times |V^R|}$, where $z_{xm} = 1$ indicates a matching of foreground regions between V_x^S and V_m^R. For computation purposes, we make $\mathbf{Z}$ into a column vector $\mathbf{z} \in \{0, 1\}^{|V^S||V^R|}$, and apply the method of [3] for solve the following problem instead:

$$\max \left(\mathbf{z}^T \mathbf{W} \mathbf{z}\right), \text{s.t.} \sum_{k=1}^{|V^R|} z_{xk} \leq 1, \quad \sum_{k=1}^{|V^I|} z_{km} \leq 1, \ \ \forall x, m. \quad (3)$$

The constraints in (3) enforces one-to-one matching between G^S to G^R. After solving (3), the locations of the foreground objects in the source image can be updated accordingly. As depicted in Figure 3, we denote the original location of objects in the source image as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, ...\mathbf{s}_N] \in \mathbb{R}^{2\times N}$, the estimated ones as $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ...\mathbf{p}_N] \in \mathbb{R}^{2\times N}$, and the final recomposed and retargeted ones as $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, ...\mathbf{q}_N] \in \mathbb{R}^{2\times N}$ (as detailed in the following subsection).

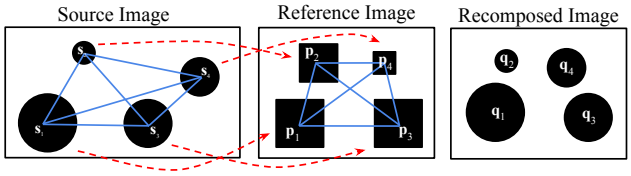


Figure 3: Graph matching for recomposition. The blue lines denote the *composition graphs*, and the red dotted lines associate the foreground objects across images. Note that **s**, **p**, and **q** denote the object locations in each image.

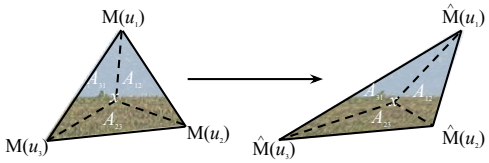


Figure 4: Illustration of mesh-based texture mapping.

2.3 Joint Recomposition and Retargeting

2.3.1 Background Mesh Warping

From Section 2.2, we estimate the new locations of the foreground objects in the source image for recomposition purposes. To complete the proposed R2P process, we propose to advance background warping for joint recomposition and retargeting. This allows us to move the foreground objects to the target locations without leaving any unfilled or distorted regions in the background.

We note that, the recent works of [2] and [10] apply image inpainting techniques for producing their recomposed outputs. Such techniques require the background regions to exhibit proper textural properties. Moreover, it is also computationally expensive to perform image inpainting (compared to our warping-based algorithm).

In [5], a uniform stretch method was considered to construct image meshes for image retargeting. Their goal was to minimize the predicted error of the recovered image with consistency in the observed boundaries and edges. Since their approach is not able to alter the composition of the input image (i.e., recomposition), we propose a novel mesh-warping algorithm for solving the tasks of recomposition and retargeting simultaneously.

To be more specific, our goal is to estimate the locations of foreground objects and the image cropping factor, which would that fulfill the following two requirements:

1. Recompose and retarget the input to fit the reference.
2. Minimize the *distortion* of the output image.

With the above two goals, our task is to determine the optimal locations Q for each foreground object and the image cropping factor $\mathbf{b}$, which are defined as follows:

$$\begin{aligned} Q &= [\mathbf{q}_1, \mathbf{q}_2, ...\mathbf{q}_N] \in \mathbb{R}^{2\times N} \\ \mathbf{b} &= [b_{top}, b_{bottom}, b_{left}, b_{right}] \in \mathbb{R}^{1\times 4}, \end{aligned} \tag{4}$$

where each b indicates the amount of the image boundary to be cropped (negative b values mean image stretching). Thus, the objective function to be optimized for joint recomposition and retargeting is:

$$[Q, \mathbf{b}] = \underset{Q,\mathbf{b}}{\operatorname{argmin}}\, D(Q, \mathbf{b})E(Q, \mathbf{b}). \tag{5}$$

Algorithm 1 Optimization for Background Mesh Warping

Sort O in decreasing order of object size.
for all $u \in M, n \in [1, N]$ **do**
  $c_n(u) = 1/distance(u, O_n)^2$.
end for
for all Possible decision of $\mathbf{b}$ **do**
  for $n = 1$ to N **do**
    for all $\mathbf{q_n}, \|\mathbf{q_n} - \mathbf{p_n}\| < r_{max}$ **do**

$$\hat{\mathbf{M}}_n(u) = \frac{c_n(u)\left[\mathbf{M}(u) + \mathbf{q}_n - \mathbf{s}_n\right] + \sum_{k=1\to n-1} c_k(u)\hat{\mathbf{M}}_k(u)}{\sum_{k=1\to n} c_k(u)}, \forall u$$

      Handle boundary constraint.
      Calculate $D(\mathbf{q_n}, \mathbf{b})$ and $E(\mathbf{q_n}, \mathbf{b})$.
    end for
    $\hat{\mathbf{q}}_n = \operatorname{argmin}_{\mathbf{q_n}} D(\mathbf{q_n}, \mathbf{b})E(\mathbf{q_n}, \mathbf{b})$
  end for
  $Q_{\mathbf{b}} = [\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, ...\hat{\mathbf{q}}_N]$
end for
Find $\hat{B} = \operatorname{argmin}_{\mathbf{b}} D(Q_{\mathbf{b}}, \mathbf{b})E(Q_{\mathbf{b}}, \mathbf{b})$
return $\hat{Q} = Q(q_{\hat{\mathbf{b}}}, \hat{\mathbf{b}})$

The first objective function D jointly solve recomposition and retargeting as follows:

$$D(Q, \mathbf{b}) = \prod_{n\in[1,N]} (1 + \alpha\|\mathbf{q}_n - \mathbf{p}_n\|) \prod_{b_i\in\mathbf{b}} (1 + \beta\|b_i\|), \tag{6}$$

To avoid remarkable changes between the original and output images, we set the constraint of the foreground object locations in the output image as $\|\mathbf{q}_n - \mathbf{p}_n\| < r_{max}$. We also limit the cropping factor $\|b_i\| < b_{max}$. In our work, parameters r_{max} and b_{max} are set as 10% and 20% of the width of the source image, respectively.

On the other hand, the objective function E suppresses image distortion, which is determined by:

$$E(Q, \mathbf{b}) = (1 + \gamma F) \sum_{L_x\in L} \sum_{(u,v)\in L_x} \left\|\hat{\mathbf{M}}(u, v) - \mathbf{M}(u, v)\right\|^2, \tag{7}$$

where $\mathbf{M}$ and $\hat{\mathbf{M}}$ represent the 2D coordinates of source and recomposed mesh points (see Figure 2d and 2e for examples), and u and v denote the indices of the corresponding mesh points. Thus, $\mathbf{M}(u, v)$ represents the vector connecting points u and v. F is the proportion of flip-over triangle in the recomposed output, L denotes the set of strong edges (see Section 2.1).

To solve (5) for producing the final output image, we calculate the distances between each mesh point and the foreground objects, and update the locations of the mesh points $\hat{M}$ (via iterating between the calculation of $\mathbf{q}$ and $\mathbf{b}$). Note that, for display purposes, we show the recomposed but uncropped output in Figure 2e (i.e., no updates of $\mathbf{b}$). As for satisfying the boundary constraints to minimize image distortion, we uniformly stretch the background mesh grids to meet the nearest image boundary. The optimization process is summarized in Algorithm 1.

2.3.2 Texture Mapping for Image Completion

Once the locations of each mesh point in $\hat{\mathbf{M}}$ are determined, the remaining task is to render the final image output. This is achieved by texture mapping. That is, for each

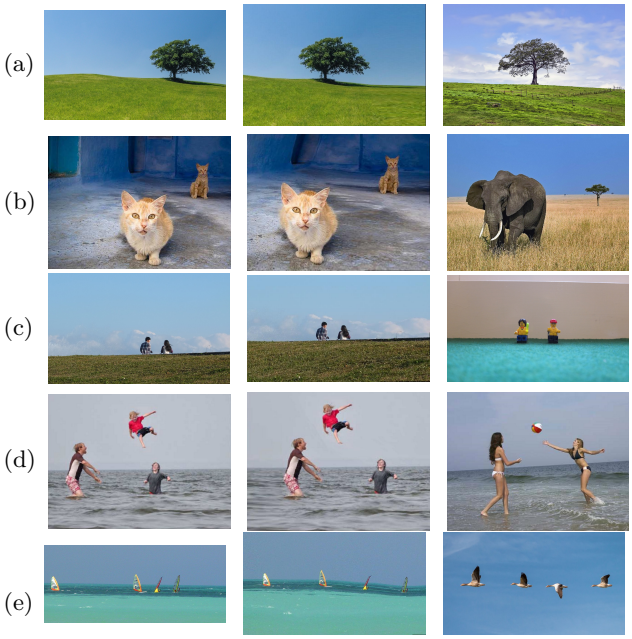


Figure 5: Examples of photography recomposition and retargeting with different foreground object numbers.

triangle in $\mathbf{M}$, we calculate the areas A_{12}, A_{23} and A_{31} for each pixel $\mathbf{x}$ inside the triangular mesh (see Figure 4 for illustration). By barycentric interpolation, the pixels in $\hat{\mathbf{M}}$ can be computed as follows:

$$\hat{\mathbf{x}} = \left[A_{23}\hat{\mathbf{M}}(u_1) + A_{31}\hat{\mathbf{M}}(u_2) + A_{12}\hat{\mathbf{M}}(u_3)\right] /(A_{23}+A_{31}+A_{12}). \quad (8)$$

After all pixels are derived, the output will be the final recomposed and retargeted image.

3. EXPERIMENTS

We consider a variety of photographic images for evaluation. All the image data are collected from either those considered in related works or the Internet[1] . Figure 5 shows the results produced by our method. It can be seen that, given different reference images, our method successfully recomposed and retargeted images with single or multiple foreground objects. For example, we were able to transfer the composition and preserve the foreground-to-background relationships (see the shadow of the cat in Figure 5b and the grass regions in Figure 5c). As noted in [2], we can only provide qualitative evaluation for the effectiveness of recomposition, since no aesthetic scores can be calculated (we do not apply any predetermined aesthetic rules).

Compared to a recent work of [2], they required complex post-processing procedures to refine the recomposed outputs to avoid possible background mismatch. Moreover, they did not address the task of retargeting, so their outputs cannot fit the references well (see Figure 6 for example).

Finally, we present results in Figure 7 to show that our method can retarget the images, so that the outputs would

[1] Fig.1-s: CC Attribution, Thomas Leuthard; Fig.5a-r: hjjanisch @Flickr; Fig.5b-s, 5d-s: cited from [10]; Others: free or CC0 or public image from Internet.

Results of [2]. Our output image

Figure 6: Comparison between [2] and our approach. Note that [2] simply performs image inpainting for recomposition, and thus its output does not fit the reference in Figure 5c.

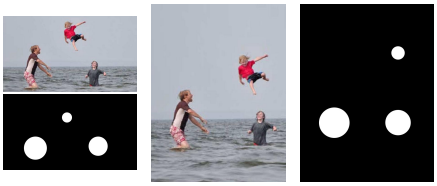


Figure 7: Retargeted outputs of Figure 5d using reference images (in black and white) with different aspect ratios.

fit the reference images with different aspect ratios. From the above experiments, we confirm the the effectiveness and robustness of our proposed method for joint recomposition and retargeting photographic images.

4. CONCLUSION

We presented an automatic framework to perform joint recomposition and retargeting of the input image for fitting the reference of interest. Unlike prior works requiring pre-collected data or predetermined aesthetics rules, our method applies graph matching to predict the locations of the foreground objects, while both foreground and background regions will be retargeted to fit the reference image. The above process is completed by solving a constrained mesh-warping based optimization problem, which minimizes both recovered errors and image distortion. Our experiments on several real-world photographic images with different foreground numbers verified the effectiveness of our method.

Acknowledgement This work is supported in part by the Ministry of Science and Technology of Taiwan via MOST104-3011-f-002-005 and MOST103-2221-E-001-021-MY2.

5. REFERENCE

[1] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *ACM Multimedia*, 2010.

[2] H.-T. Chang, Y.-C. F. Wang, and M.-S. Chen. Transfer in photography composition. In *ACM Multimedia*, 2014.

[3] T. Cour et al. Balanced graph matching. In *NIPS*. 2007.

[4] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *IEEE CVPR*, 2010.

[5] Y. Guo et al. Image retargeting using mesh parametrization. *IEEE Trans. Multimedia*, 2009.

[6] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or. Optimizing photo composition. *Computer Graphics Forum*, 2010.

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

[8] A. Santella et al. Gaze-based interaction for semi-automatic photo cropping. In *SIGCHI*, 2006.

[9] L.-K. Wong and K.-L. Wong. Enhancing visual dominance by semantics-preserving image recomposition. In *ACM Multimedia*, 2012.

[10] F.-L. Zhang, M. Wang, and S.-M. Hu. Aesthetic image enhancement by dependence-aware object recomposition. *IEEE Trans. Multimedia*, 2013.