Content-Aware Image Retargeting Incorporated with Letterboxing

Kazu MISHIBA\(^{(a)}\), Member, Yuji OYAMADA\(^1\), Nonmember, and Katsuya KONDO\(^{1}\), Member

SUMMARY Conventional image retargeting methods fail to avoid distortion in the case where visually important regions are distributed all over the image. To reduce distortions, this paper proposes a novel image retargeting method that incorporates letterboxing into an image warping framework. Letterboxing has the advantage of producing results without distortion or content loss although being unable to use the entire display area. Therefore, it is preferable to combine a retargeting method with a letterboxing operator when displaying images in full screen. Experimental results show that the proposed method is superior to conventional methods in terms of visual quality measured by an objective metric.

key words: content-aware image retargeting, image resize, image warping, letterboxing

1. Introduction

As the diversity of consumer display devices increases (e.g., smart phones, notebooks, tablets, and HDTV), it is increasingly important that visual media display appropriately on these devices. Because the same image needs to be displayed with different resolutions on different devices, image resizing plays an important role in appropriate display. The classical resizing methods of homogeneous scaling, cropping, and letterboxing\(^*\) have the drawbacks of distorting the entire image, discarding important parts of the image, and preventing the use of the entire display area. Content-aware image retargeting\([1, 2]\) has recently been developed to overcome these limitations. The key concept of these techniques is to resize images to an arbitrary resolution while protecting visually important regions from distortion. To achieve it, a content-aware image retargeting method deformed unimportant regions because their distortion is less noticeable.

Several retargeting methods have been proposed, including seam carving\([3–7]\) and image warping\([8–13]\). These methods, however, fail to avoid distortion in the case where visually important regions are distributed all over the image.

To solve this problem, this paper proposes a retargeting method that combines with cropping and letterboxing operators in an image warping framework. Letterboxing has the disadvantage of being unable to use the entire display area but the advantage of producing results without distortion or content loss. Images viewed on consumer devices are often displayed in full screen. Therefore, it is preferable to combine a retargeting method with a letterboxing operator.

2. Related Work

Content-aware image retargeting changes the size of a source image while trying to protect visually important regions from distortion. Visual importance can be calculated by many methods: image gradients, a saliency measure\([14–16]\), face detection, and combinations or modified version of these methods. These methods create an importance map, which stores the relative importance value for each pixel. The core goal of retargeting methods is to change the size of an image while limiting distortion in important regions.

Seam carving\([3]\) and its expansions\([4–7]\) are one commonly used approach for image retargeting. Seam carving changes the size of a source image by iteratively carving out or inserting paths of pixels, which is called seam. A seam is selected to pass through less important regions to keep important regions unchanged. In other words, seam carving deforms images while keeping important regions as rigid as possible. In the seam carving approach, the position of contents in an image changes along one direction, e.g., the position of contents changes only horizontal direction when the width of an image changes.

Warping methods are another commonly used approach for image retargeting. They can change the position of contents along vertical and horizontal directions when the width of an image changes. They place a mesh onto an image and then deform the mesh by solving an optimization problem. Many methods use a regular grid mesh and few methods\([10]\) use an irregular triangle mesh. One of earlier studies of warping methods\([13]\) deforms a mesh while applying a similarity or a rigid transformation to user-specified regions. A rigid transformation keeps the sizes of regions and a similarity transformation keeps the aspect ratios of regions. The key idea of recently developing methods is to apply these shape-protecting transformations to visually important regions while allowing arbitrary deformations to unimportant regions. This nonhomogeneous deformation

\(^*\)Letterboxing/pillarboxing adds black bars to the top and bottom or the left and right of an image. For simplicity, the discussion in this paper is limited to the case of letterboxing, but can be applied to pillarboxing.
approach produces distortions, such as grid line bending and self-intersection of a mesh. Several methods have been proposed to prevent these distortions. Wang et al. [9] introduced an energy term which penalizes grid edge orientations to prevent excessive grid line bending. Self-intersection introduces image discontinuity. In linear least-squares optimization, it is difficult to prevent self-intersection. Thus, several methods [11], [13] fail to prevent self-intersection. In non-linear optimization, several methods prevent self-intersections by enforcing the flipping edge to have zero length during iterations [9], penalizing reversal of triangle orientation with sign function [10], and constraining the minimal allowed size of grid cells [8].

To produce satisfactory results with fewer artifacts, Rubinstein et al. [17] proposed a framework to combine several retargeting operators. In their approach, combination of several retargeting operators results in the violation of reduction of the size of important objects, combining with a letterboxing operator reduces the displayable area, resulting in distortion caused by other operators. The proposed method in this paper have modified energy functions for retargeting to improve quality of retargeted images.

Conduction of subjective tests to evaluate retargeting results is expensive and time-consuming. Thus visual quality assessment for retargeted images is important. In Rubinstein’s study [19], the authors used some objective distance metrics, such as bi-directional similarity [20], color layout descriptor [21], BDW [17], edge histogram [22], SIFT-flow [23], and Earth-Mover’s Distance [24], then estimated how well the objective metrics agree with the users’ subjective preferences. Ma et al. [25] conducted similar experiments. These two studies showed that the performances of the objective quality metrics for retargeted images are still not good enough. Hsu et al. [26] proposed an objective metric based on perceptual geometric distortion and information loss. The geometric distortion of a retargeted image is measured based on the local variance of SIFT-flow vector fields [23] of the image. The information loss in the retargeted image is measured based on the saliency map. Their experimental results of subjective tests showed that their metric exhibits the good consistency with the subjective rankings. In this paper, evaluation of visual quality of retargeted images is conducted by using the metric proposed by Hsu et al. [26].

3. Proposed Retargeting Method

3.1 Problem Formulation

For simplicity, the following discussion is limited to the case of reducing the image size in the horizontal direction.

The proposed method combines a rigid and similarity transformations, and a cropping and letterboxing operators in an image warping framework. The effect of these operators depends on images due to a content-aware manner. In the proposed warping method, a grid mesh is projected onto the image and then deformed to a target size to satisfy desired conditions as much as possible by solving an energy optimization problem. An initial grid mesh consists of horizontal $\alpha$ lines and vertical $\beta$ lines. Many conventional warping methods move the $\alpha \beta$ vertices of the grid mesh in two dimensions. In contrast, the proposed method moves the $\alpha + \beta$ grid lines in one dimension. In other words, the intervals of the neighboring lines are changed for retargeting. Compared with methods in which the vertices move in two dimensions, this approach provides the following two advantages. The first advantage is faster computation because of the reduction of the unknowns of the optimization prob-
lem from $2\alpha \beta$ to $\alpha + \beta$. The second advantage is free from bending of grid lines. Whereas some researches [9], [12] require an energy term which penalizes grid edge orientations to prevent excessive grid line bending, the proposed approach is free from bending of grid lines in principle.

The initial grid mesh $G = \{V, U, F\}$ is represented as a set containing three components. $V = \{X, Y : X = \{x_i\}_{i=1}^n, Y = \{y_j\}_{j=1}^m\}$ is a set of grid lines where $x_i > x_{i+1}$, $y_{j+1} > y_j$, $x_1 = 0$, $x_n = W$, $y_1 = 0$ and $y_m = H$. Here $W$ and $H$ are the width and the height of a source image, respectively. $U = \{\delta X, \delta Y : \delta X = \{\delta x_i\}_{i=1}^{n-1}, \delta Y = \{\delta y_j\}_{j=1}^{m-1}\}$ is a set of grid sections between neighboring lines where $\delta x_i = x_{i+1} - x_i$ and $\delta y_j = y_{j+1} - y_j$. $F = \{f = (\delta x_i, \delta y_j) : \delta x_i \in \delta X, \delta y_j \in \delta Y\}$ is a set of faces where $f = (\delta x_i, \delta y_j)$ is a face on which $\delta x_i$ and $\delta y_j$ intersect. Let us denote an element after retargeting by adding a prime mark, e.g., $x'_i$ indicates a grid line after retargeting. The transformation of an image of size $W \times H$ to $W' \times H'$ is a problem of finding optimal grid lines $V'$ by solving an optimization problem. Here, we assume $W, H, W', H' > 0$. In the optimization problem, retargeting operators are expressed as soft and hard constraints. The soft constraints are defined as quadratic functions $E$. $E$ increases less as the deformed mesh better reflects the desired condition of a corresponding operator. The hard constraints are defined as linear equality and linear inequality constraints. The optimal grid mesh transformation, which satisfies these constraints as much as possible, is expressed as

$$V''_{opt} = \arg \min_{V''} \sum_{z \in Z} \lambda_z E_z(V'')$$

subject to hard constraints, where $Z$ is a set of soft constraints, $E_z$ is a function for soft constraint $z$ and $\lambda_z$ is a weight parameter for $E_z$. Using a larger $\lambda_z$ enhances the effect of an operator corresponding to $z$. Equation (1) is a linear least-squares problem with linear inequality and equality constraints in the proposed method and has a unique solution (see Appendix). It can be solved using an active set method and other efficient methods. Constraints corresponding to operators used in the proposed method are described in Sect. 3.3.

3.2 Retargeting Process

Figure 2 is an overview of the proposed retargeting method. It first projects an initial mesh onto an image and calculates an importance map, then obtains a crop range for a cropping operator. Next, the initial mesh is deformed by solving (1). Finally, the original image is deformed according to mesh deformation to obtain a retargeted image.

The importance map $S \in \mathbb{R}^{W \times H}$ is computed using the saliency detection proposed by Goferman et al. [27]. This method has suitable characteristics for retargeting because it detects the image regions that represent the scene, which differs from methods whose goal is to detect the dominant object. Conventional retargeting methods such as [12], [28] use [27] for computation of an importance map. Importance values are normalized in the range of $[0, 1]$. An importance value at a pixel $(x, y)$ is expressed as $s(x, y)$.

A crop range is calculated based on the image importance $S$. It is preferable that high importance regions are not cropped. The proposed method defines high-importance regions as regions satisfying $S > t_h$, where $t_h$ is a threshold parameter, $t_h = 0.2$ in this paper. A minimum rectangle including all high-importance regions will be referred to as the cropping window. The proposed retargeting method trims the right and left sides of an image, not the top and bottom sides. Let the nearest grid lines outside a cropping window on the left and right side be $\mu_l$ and $\mu_r$. A crop range is defined as the ranges from the left-side boundary to $\mu_l$ and from $\mu_r$ to the right-side boundary. Figure 3(b) presents an example of an importance map and the automatic cropping window.

3.3 Constraint Definition

The proposed method combines a rigid and similarity transformations, and a cropping, and letterboxing operators in an image warping framework. A rigid and similarity transformations, both of which are often used in conventional warping methods, protect the size and the aspect ratio of important regions respectively while deforming unimportant regions. A cropping operator efficiently discards unimportant regions on the side of an image. Discarding these regions by cropping is better than deformation by a rigid and similarity transformations because cropping is free from artifact in principle. A letterboxing operator virtually protects the aspect ratio of all the contents of an image. This operator avoids discarding and deformation of important regions,
which are often caused by other operators. In addition to above operators, the proposed method uses the following two constraints. One is to avoid excessive mesh shrinkage and the other is a boundary condition.

The rest of this subsection describes each constraint in (1) corresponding to each operator and condition.

**Cropping.** The constraint for cropping is described by a soft and hard constraints. The hard constraint, which defines the range to avoid cropping, is expressed as follows:

\[
\begin{align*}
  x'_i & \geq 0 \quad \text{if} \quad x_i \geq \mu_l, \\
  x'_i & \leq W' \quad \text{if} \quad x_i \leq \mu_r, \\
\end{align*}
\]

(2)

Because the hard constraint imposes no penalty, the cropping operator is preferentially applied over the other operators. This approach is similar to Wang’s method [29]. However, the method fails to define the range to be cropped if the target width \(W'\) is greater than the width of the cropping window, i.e., \(W' > \mu_r - \mu_l\). The proposed method adds a soft constraint to appropriately decide how to crop. The constraint imposes a retargeted image to keep the relative horizontal position of an approximate importance centroid in the image, defined by

\[
x_c = \frac{1}{s_{all}} \sum_{\delta x_i \in \Omega X} s_{section}(\delta x_i) \left( x_i + \frac{\delta x_i}{2} \right),
\]

(3)

where

\[
s_{all} = \int_{\delta X} \int_{\delta Y} s(x,y) dxdy,
\]

(4)

\[
s_{section}(\delta x_i) = \int_{\delta X} \int_{\delta x_i} s(x,y) dxdy
\]

(5)

are the sum of the importance values across the entire image and the sum of the importance values in a section \(\delta x_i\), respectively. The soft constraint is expressed as follows:

\[
E_c(V') = \left( \frac{x_c}{W} - \frac{x'_c}{W'} \right)^2.
\]

(6)

\(E_c\) increases more with an increasing difference between the relative positions of an approximate importance centroid before and after retargeting. Figure 4 presents an example of the effect of the cropping operator. The retargeted result of Fig. 4 (c) is uniquely determined and is comparatively keep the relative position of the main object.

**Rigid transformation.** The constraint for a rigid transformation that preserves the size of important regions as much as possible is described by a soft constraint. The condition that completely satisfies the constraint is that the width and height of regions remain invariant. Thus, the soft constraint is expressed as follows:

\[
E_r(V') = \sum_{\delta x_i \in \Omega X} s_{section}(\delta x_i) \left( 1 - \frac{\delta x'_i}{\delta x_i} \right)^2
\]

is the sum of the importance values in a section \(\delta y_j\).

**Similarity transformation.** The constraint for a similarity transformation that preserves the aspect ratio of important regions as much as possible is described by a soft constraint. The condition that completely satisfies the constraint is that the width and height of a region are scaled at the same rate. From this observation, the soft constraint is expressed as follows:

\[
E_s(V') = \sum_{f=(\delta x_i, \delta y_j) \in F} s_f \left( \frac{\delta x'_i}{\delta x_i} - \frac{\delta y'_j}{\delta y_j} \right)^2
\]

where

\[
s_f = \frac{1}{|f|} \int_{\delta y_j} \int_{\delta x_i} s(x,y) dxdy
\]

is the average of the importance values on face \(f = (\delta x_i, \delta y_j)\) and \(|f|\) is the area of \(f\). \(E_s\) increases more with an increasing difference in the scaling rates of the width and height of a face.

**Letterboxing.** The constraint for letterboxing is described by a soft constraint. The soft constraint is expressed as follows:

\[
E_l(V') = w \left( \frac{l}{H} \right)^2
\]

where

\[
l = H' - (y'_p - y'_l)
\]

corresponds to the height of black bars to be added for letterboxing and \(w\) is a weight parameter. \(E_l\) increases more with an increasing the height of the black pixels. Important regions are likely to be deformed by the rigid and similarity transformations when distributed widely over the image. Thus, to reduce distortions caused by these transformations,
it is preferable to increase the height of black bars when important regions are more widely distributed. To achieve it, when important regions are more widely distributed, weight \( w \) is set to a smaller value, which can encourage to increase the height of black bars due to reduction of \( E_i \). The proposed method expresses the distribution of important regions by using the standard deviation of the importance map. Let \( \sigma_S \) and \( \bar{S} \) be the standard deviation of the importance map \( S \) and the average of \( S \). Smaller \( \sigma_S \) means wider distribution of important regions. The proposed method uses the following \( w \):

\[
w = \frac{\sigma_S}{\bar{S}}.
\]  

Constraint to avoid excessive mesh shrinkage. The condition that avoids excessive mesh shrinkage is that the length between neighboring grid lines is not less than a certain length. From this observation, the hard constraint to avoid mesh shrinkage is expressed as follows:

\[
\delta x_i' \geq \mu_m \delta x_i \frac{W'}{W}, \quad \delta y_j' \geq \mu_m \delta y_j \frac{H'}{H},
\]  

where \( \mu_m = [0, 1] \) is an adjustment parameter, \( \mu_m = 0.2 \) in this paper. Figure 5 presents an example of preventing excessive mesh shrinkage. The hard constraint also prevents self-intersection of a mesh. While some methods [11], [13] fail and other methods [8]–[10] require the update of weight parameters or energy functions during an energy optimization process to prevent self-intersection, the proposed method guarantees prevention of self-intersection and does not require such updates.

Boundary condition. Warping methods require a boundary condition. The boundary condition of the proposed method differs from those of typical warping methods in which the location of the grid lines on the boundary is restricted to the target image boundary. A retargeted image must be at the vertical center of the display to achieve letterboxing display. This condition is satisfied by the following constraint:

\[
y_1' + y_b' = H'.
\]  

In the proposed method, the vertical grid lines at the boundary are allowed to be on and outside the target image boundary but are not allowed to be inside the boundary because such a configuration would result in pillarboxing or windowboxing. Thus, the proposed method uses the following hard constraint:

\[
x_1' \leq 0, \quad x_n' \geq W'.
\]  

4. Experimental Results

The proposed method have been implemented in MATLAB with Optimization Toolbox to solve the optimization problem. Our experiments were conducted on a PC with an Intel i7 3.40 GHz CPU and 12 GB RAM. The experiments used a 40 × 40 grid mesh, i.e., \( \alpha = 40 \) and \( \beta = 40 \), divided by equal intervals. Weight parameters for the cropping, rigid transformation, similarity transformation, and letterboxing were set to \( \lambda_c = 1, \lambda_r = 1, \lambda_s = 500, \) and \( \lambda_l = 500 \), respectively.

The proposed method was evaluated using 80 test images containing various types of scenes, such as people, clear foreground objects, natural scenery, and geometric structures. Part of the image sequence is shown in Fig. 6. The width of these images is 768 or 800 pixels and was reduced by half in a retargeting process. The proposed method was compared with the two conventional warping methods, optimized scale-and-stretch resizing (SS) [9] and patch-based image warping (PB) [12]. In addition, the proposed method was also compared with multi-operator retargeting (MO) [17] which uses three operators (cropping, scaling and seam carving). All the methods excluding SS computed a saliency map proposed by Goferman et al. [27]. In SS, a saliency map was computed by the method written

\[
\text{Table 1} \quad \text{The number of ranks that the five methods earn through all the test images. A method which has the smallest retargeting quality index } q_{\text{retarget}} \text{ among five methods takes the first rank for each image.}
\]

| Method   | 1st | 2nd | 3rd | 4th | 5th | Average \( q_{\text{retarget}} \) |
|----------|-----|-----|-----|-----|-----|-------------------------------|
| Letterboxing | 5   | 15  | 18  | 16  | 26  | 0.377                         |
| SS [9]    | 4   | 13  | 19  | 21  | 23  | 0.394                         |
| PB [12]   | 6   | 9   | 14  | 29  | 22  | 0.393                         |
| MO [17]   | 32  | 16  | 19  | 6   | 7   | 0.352                         |
| Proposed  | 33  | 27  | 10  | 8   | 2   | 0.337                         |

Fig. 6 Part of test images
in Wang’s paper [9] because SS was performed with the executable file created by the authors. To show the effectiveness of combining letterboxing into content-aware image retargeting, the test images were also resized using only letterboxing, which is not content-aware image retargeting. Visual quality assessment for retargeted images was conducted by using the objective metric proposed in [26], named retargeting quality index $q_{\text{resize}} \in [0, 1]$. This index is calculated based on perceptual geometric distortion and information loss. A smaller index means better visual quality.

Table 1 shows the summary of the comparison results. Although a letterboxing operator produces retargeted results without distortion or content loss, it drastically reduces the size of contents. The objective metric penalizes such reduction, resulting in worse $q_{\text{resize}}$. The MO approach and the proposed method often exhibit better $q_{\text{resize}}$ than other methods.

![Fig. 7 Retargeting results and $q_{\text{resize}}$. (a) Original images. (b) Letterboxing. (c) SS [9]. (d) PB [12]. (e) MO [17]. (f) Proposed.](image-url)
Fig. 8 Retargeting results of the proposed method with a grid mesh (red) and a cropping window (green).

Fig. 9 Retargeting results with different grid resolutions. (a) $\alpha = \beta = 20$. (b) $\alpha = \beta = 30$. (c) $\alpha = \beta = 40$.

methods. Although the MO approach can produce plausible retargeting results, its computational cost is very high. On average, for a $800 \times 600$ image, the computation time for resizing (excluding for saliency calculation) is 0.34 seconds in the proposed method, 5.67 seconds in the PB approach, and more than 40 minutes in the MO approach. Exhaustive search in the MO approach to find the optimal combination of operators requires high computational cost. In contrast, the proposed method requires no exhaustive search and only solves a linear least-squares problem with linear inequality and equality constraints to produce resized images. In addition, the proposed method achieved faster computation than the PB approach, which uses a mesh transformation. This is because the proposed method has fewer unknowns in the optimization problem than the PB approach due to the gridline-based mesh transformation.

The visual comparisons are shown in Fig. 7. The retargeting results of the proposed method with a mesh and a cropping window are shown in Fig. 8. The original image in Fig. 7 (i) contains clear foreground objects. In the conventional warping methods SS and PB, the shapes of foreground objects are distorted. While preserving the shapes, the MO approach significantly removes backgrounds, producing artifacts. It is observed that the proposed method produces a retargeted image with less distortion and artifacts among all content aware retargeting methods due to combination with a letterboxing operator. Although the proposed method reduces the displayable area due to black bars, the height of the bars in the proposed method is shorter than in using only letterboxing. Thus the proposed method avoids overshrinkage of contents. Figure 7 (ii) and (iii) show a clear difference between the conventional methods and the proposed method. The original images include visually important regions, in this case human figures, over the images. Whereas the conventional methods largely deformed these regions, the proposed method reduced these deformation due to the letterboxing operator. In some cases, the proposed method exhibits worsen $q_{\text{resize}}$ because of over-cropping. Although the result of Fig. 7 (iv) of the proposed method seems less artifacts compared with other retargeting methods, the proposed method is inferior in $q_{\text{resize}}$ to the MO approach. Because the background of the original image includes some contents, the loss of them worsens $q_{\text{resize}}$. In Fig. 7 (v), a part of the face of the girl on the right is trimmed out because of an undesirable cropping window, as shown in Fig. 8 (v). Although the proposed method achieved better preservation of the aspect ratio of the girls’ faces than other methods, partial loss of important contents results in an undesirable retargeting result. Since a cropping window largely depends on a saliency measure method, development of better saliency measure methods may improve results of the proposed method.

Finally, we discuss the resolution of a grid mesh in our method. Figure 9 shows the results of retargeting of Fig. 3 (a) with different grid resolutions. As can be seen from the figure, the grid resolution affects retargeting results. This is because the grid resolution affects the relative value of $E_s$. For example, $E_s$ is independent and $E_t$ is dependent on the grid resolution. Therefore the relative effect of each operator is affected by the grid resolution.

5. Conclusion

This paper proposed a novel image retargeting method that combines a rigid and similarity transformations, and a cropping, and letterboxing operators in an image warping framework. Combining a letterboxing operator effectively re-

---

Because SS was performed with the executable file, which requires user interaction, we did not measure the computation time.
duces distortions which occur when visually important regions are distributed all over the image. Compared with Rubinstein’s method [17], which combines multiple operators sequentially, the proposed method combines them in a warping framework. It avoids exhaustive search and requires lower computational time.

Acknowledgments

We thank the following Flickr (http://www.flickr.com/) users for Creative Commons imagery: gwenael.piaser (Two Little Tourists), Noël Zia Lee (Bird), Tim Wilson (Racing), Paraflyer (three parrots), Aaron Landry (Flower Cupcakes), Joe Shlabotnik (playground), melississippi (Snow Mountain), Paraflyer (One Parrot), B Milly (soccer), Michael Galkovsky (Farm), jasonhill (Two Girls), makelessnoise (Close Friends), Stuck in Customs (Car), moriza (Chicken Paraflyer (three parrots), Aaron Landry (Flower Cupcakes), We thank the following Flickr (http://www.flickr.com/) users for Creative Commons imagery: gwenael.piaser (Two Little Tourists), Noël Zia Lee (Bird), Tim Wilson (Racing), Paraflyer (three parrots), Aaron Landry (Flower Cupcakes), Joe Shlabotnik (playground), melississippi (Snow Mountain), Paraflyer (One Parrot), B Milly (soccer), Michael Galkovsky (Farm), jasonhill (Two Girls), makelessnoise (Close Friends), Stuck in Customs (Car), moriza (Chicken Paraflyer (three parrots), Aaron Landry (Flower Cupcakes),

References

[1] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand, “A survey of image retargeting techniques,” Proc. SPIE Applications of Digital Image Processing XXXIII, 2010.

[2] A. Shamir and O. Sorkine, “Visual media retargeting,” SIGGRAPH ASIA Courses, New York, NY, USA, pp.1–13, ACM, 2009.

[3] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” ACM Trans. Graph., vol.26, no.3, p.10, 2007.

[4] M. Rubinstein, A. Shamir, and S. Avidan, “Improved seamless carving for video retargeting,” ACM Trans. Graph., vol.27, no.3, pp.1–9, 2008.

[5] S. Cho, H. Choi, Y. Matushita, and S. Lee, “Image retargeting using importance diffusion,” Proc. IEEE Int. Conf. Image Processing, pp.977–980, 2009.

[6] K. Mishiba and M. Ikehara, “Seam merging for image resizing with structure preservation,” Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp.1001–1004, 2011.

[7] K. Mishiba and M. Ikehara, “Image resizing using improved seam merging,” Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp.1261–1264, IEEE, 2012.

[8] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, “A system for retargeting of streaming video,” ACM Trans. Graph., vol.28, no.5, pp.126:1–126:10, Dec. 2009.

[9] Y-S. Wang, C-L. Tai, O. Sorkine, and T-Y. Lee, “Optimized scale-and-stretch for image resizing,” ACM Trans. Graph., vol.27, no.5, pp.118:1–118:8, 2008.

[10] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, “Image retargeting using mesh parametrization,” IEEE Trans. Multim., vol.11, no.5, pp.856–867, 2009.

[11] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R.R. Martin, “A shape-preserving approach to image resizing,” Comput. Graph. Forum, vol.28, no.7, pp.1897–1906, 2009.

[12] S-S. Lin, I-C. Yeh, C-H. Lin, and T-Y. Lee, “Patch-based image warping for content-aware retargeting,” IEEE Trans. Multimedia, vol.15, no.2, pp.359–368, 2013.

[13] R. Gal, O. Sorkine, and D. Cohen-Or, “Feature-aware texturing,” Proc. Eurographics Symposium on Rendering, pp.297–303, 2006.

[14] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.1–8, 2007.

[15] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.20, no.11, pp.1254–1259, 1998.

[16] R. Achanta, S. Hemami, E. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.1597–1604, 2009.

[17] M. Rubinstein, A. Shamir, and S. Avidan, “Multi-operator media retargeting,” ACM Trans. Graph., vol.28, no.3, pp.1–11, 2009.

Appendix: Proof of Uniqueness of Mesh Transformation

The optimization problem in our proposed mesh transformation is

\[ V'_{opt} = \arg \min_{V'} E(V') \quad \text{s.t.} \quad (2), (14), (15), (16) \]  

where

\[ E(V') = \lambda_v E_v(V') + \lambda_s E_s(V') + \lambda_c E_c(V') + \lambda_t E_t(V'). \]  

We assume the case where \( s_{\text{section}}(\cdot), \sigma_s, \lambda_s > 0 \).

Given \( \mathbf{v}' = [x', y'] \in \mathbb{R}^2 \) where \( x' = [x'_1, \ldots, x'_d] \in \mathbb{R}^d, y' = [y'_1, \ldots, y'_d] \in \mathbb{R}^d \) and \( \gamma = \alpha + \beta \), the problem
can be written in matrix notation as the minimization of the quadratic function with linear constraints:

\[ v_{opt} = \arg \min_{v'} \nu^T P v' + q^T v' + r, \quad \text{(A-3)} \]

subject to \( A v' \leq b, \quad \text{(A-4)} \)

\[ C v' = d, \quad \text{(A-5)} \]

with \( P \in \mathbb{R}^{m \times m}, q \in \mathbb{R}^m, r \in \mathbb{R}, A \in \mathbb{R}^{m \times m}, b \in \mathbb{R}^m, C \in \mathbb{R}^{1 \times m}, d \in \mathbb{R} \) and \( m \geq 2x + \beta + 2 \). Here \( m \) depends on a crop range as shown in (2).

**Lemma 1** The problem (A-1) is feasible.

**Proof:** For \( x_i' = x_i^W + y_i^W \) and \( y_i' = y_i^W \), we have \( x_i' = 0, x_i'' = W', y_i'' = 0, y_i' = H', \delta x_i' = x_i' - x_i'' = (x_i'' - x_i)W = \delta x_i, y_i' = H' = \mu_2 \delta x_i W + \delta y \geq \mu_2 \delta y \). So we clearly see that \( x_i' \) and \( y_i' \) satisfy (2), (14), (15) and (16). Since \( v' = [x_1', y_1']^T \) is a feasible point for the problem (A-1), it is feasible.

**Lemma 2** \( P \) in (A-3) is positive definite.

**Proof:** Let \( \Gamma(v) \) be a function that leaves only quadratic terms in the variables \( x \), e.g., \( \Gamma(E(V')) = v^T P v' \). An approximate importance centroid after retargeting can be expressed as

\[ x_i' = \frac{1}{s_{all}} \sum_{x_i \in X} s_{section}(\delta x_i)(x_i' + \delta x_i') \]

\[ = \sum_i t_i x_i', \quad \text{(A-6)} \]

where \( t_i > 0 \) is a scalar. Then we deduce

\[ \Gamma(E_c(V')) = \frac{1}{W^2} \left( \sum_i t_i x_i' \right)^2 \geq 0, \quad \text{(A-8)} \]

\[ \Gamma(E_c(V')) = \sum_{\delta x_i \in \delta X} s_{section}(\delta x_i) \left( \frac{x_i' + \delta x_i'}{\delta x_i} \right)^2 \]

\[ + \sum_{\delta y_j \in \delta Y} s_{section}(\delta y_j) \left( \frac{y_j' + \delta y_j'}{\delta y_j} \right)^2 \geq 0, \quad \text{(A-9)} \]

\[ \Gamma(E_c(V')) = \sum_{f = \delta x_i, \delta y_j \in F} s_f \left( \frac{x_i' + \delta x_i}{\delta x_i} - \frac{y_j' + \delta y_j}{\delta y_j} \right)^2 \geq 0, \quad \text{(A-10)} \]

and from (11), (12) and (15)

\[ \Gamma(E_c(V')) = \frac{4W}{H^2} y_i'^2 \geq 0. \quad \text{(A-11)} \]

Since \( \Gamma(E(V')) = \lambda_i \Gamma_e(E_i(V')) + \lambda_i \Gamma_e(E_i(V')) + \lambda_i \Gamma_e(E_i(V')) + \lambda_i \Gamma_e(E_i(V')) = \nu^T P \nu', \nu^T P \nu' > 0 \) for all \( \nu' \neq 0 \) where \( \nu \) is a zero vector. Hence, \( P \) in (A-3) is positive definite.

**Proposition 1** The problem (A-1) has a unique solution.

**Proof:** Since the problem (A-1) is feasible from Lemma 1 and \( E(V') \) is a strictly convex function from Lemma 2, it has a unique solution.