Learning to Resize Image

Qi Wang, Yuan Yuan*

Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

Abstract

Content-aware image resizing has been a promising theme in the communities of image processing and computer vision. To the best of our knowledge, most existing methods for image resizing are unsupervised. These unsupervised methods may either fail to protect the interesting regions or cause distortion of the image structure. This paper presents a novel learning based method for seam carving by incorporating the learned boundary of the important content. Specifically, a novel boundary model of the *region of interest* (ROI) is learned on a set of training images at first. Then the boundary of an input image is utilized as a key prior in performing seam carving to obtain the target image. The proposed method for image resizing can generate much less seams cutting through the ROI compared with previous efforts toward the same goal. Thus, the desirable regions can be preserved in the target image and the structural consistency of the input image is naturally maintained. Experiments on two publicly available data sets demonstrate the effectiveness of the proposed method.

Keywords: image processing, machine learning, image resizing, seam carving, boundary, saliency

1. Introduction

With the rapid development of multimedia technology, numerous digital images are frequently used in people's daily life. To better share and exchange

---

*Corresponding author.
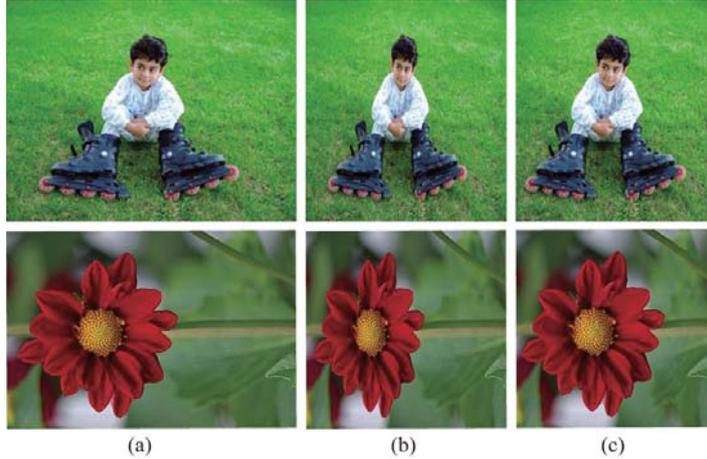*Email addresses:* `crabwq@opt.ac.cn` (Qi Wang), `yuany@opt.ac.cn` (Yuan Yuan)

Figure 1: Examples for image resizing. (a) Original input images. (b) Results of traditional nonuniform scaling. (c) Results of the proposed method. The proposed method maintains the ROI while the traditional nonuniform scaling distorts it.

information, a wide variety of display devices are developed, such as computer monitors, TV screens, and mobile phones. However, different display devices have distinct sizes and aspect ratios. If the examined image is not the same with the display device in aspect ratio, it has to be distorted to adapt to the device. This will unsatisfactorily leads to a poor visual experience. To avoid this undesirable effect, the images are expected to be automatically adapted for different display devices, which is referred to as *image resizing*. Image resizing [1] has been a promising theme in computer vision. The main goal of image resizing is to adaptively resize the images for optimal display under different conditions [2], [3]. Fig.1 shows some examples of image resizing.

Studies in psychology and cognition have found that human visual system can quickly focus on one or several interesting contents of an image at first glance. These contents are generally called *region of interest* (ROI) [4],[5],[6],[7]. The ROI and the background of an image constitute the image structure. Effective image resizing techniques should preserve the ROI as much as possible and reduce the distortion of the image structure to maintain the harmony of the image, which is called *content-aware* [8]. In order to protect the ROI in image resizing, some guiding information, which is able to distinguish the ROI from background, should be extracted. Traditional methods utilize image clues such as color and brightness contrast to infer the ROI. However, these low-level information is far less expressive for the

ROI. The resulting resized images are usually distorted. In order to improve this situation, a novel method based on machine learning is developed by incorporating a high-level boundary feature of the ROI as prior information. Fig.1 shows two typical examples of the proposed method compared with traditional nonuniform scaling. Naturally, in the two images, the boy and the flower usually attract people's eyes at first glance. The proposed method can well protect these two objects, while the traditional nonuniform scaling causes distortion of them.

The main contribution of this paper is a novel framework for image resizing by adopting supervised learning of high-level image information. We believe the incorporation of learned high-level clue is the key point to greatly improve the performance of seam carving. For this purpose, boundary is firstly taken as an example to illustrate the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 briefly reviews the related work. Section 3 presents the proposed resizing method incorporated with detected boundary information. To verify the effectiveness and robustness of the proposed method, the experimental results are shown in Section 4. Discussion is presented in Section 5 and conclusion is made in Section 6.

2. Related work

Traditional straightforward methods for image resizing, such as *scaling* and *cropping*, cannot generate satisfactory results. This is because they only consider the constraint of display space, but ignore the image content. To be specific, scaling may lead to distortions and produce artifacts when the aspect ratio of the image changes, while cropping may remove interesting areas in the result owing to an unjustified criterion. To overcome these limitations, a number of content-aware image resizing methods have been developed to preserve the ROI and avoid distortion when modifying the image's size and aspect ratio. According to the mechanism and the used methodology, these methods can be roughly classified into four categories. They are based on *content-aware cropping*, *segmentation*, *warping*, and *seam carving*, respectively.

Content-aware cropping methods [9], [10], [11], [12] are important improvements of the original cropping method. There are two main steps for this kind of methods, detecting the salient portions and cropping them by

fixed windows. However, these methods can only be applied to specific images (containing human faces or other priorly defined objects) instead of natural images, which limits their applicability. *Segmentation based methods* [9], [13], [14], [15] present a way to address the image resizing problem with segmentation techniques. The ROI is segmented from an image at first. Then the background is resized to the desired size. In the end, the segmented ROI is inpainted to the renewal background. Nevertheless, these methods rely on accurate segmentation of ROI. The inaccurate segmentation may lead to distortions of ROI. *Warping-based methods* [16], [17], [18], [19] seek to find a warping function mapping the original image grid to the target image. The local distortions of important areas are constrained to be as small as possible, while unimportant regions are allowed to distort more. However, this type of methods strongly depend on the definition of the warping function, as well as the corresponding parameter selection, which is often difficult to be satisfied in practice because different images have distinct warping requirement. *Seam carving based methods* try to obtain satisfactory resizing result by removing seams of minimal importance. A seam is defined as a path of 8-connected pixels that contains only one pixel per row or column in Avidan and Shamir [20]. The image resizing problem is actually an energy optimization problem and the dynamic programming is utilized to gracefully remove unimportant (low energy) seams in the image. After that, an improved seam carving method is developed in the work of Rubinstein *et al.* [21] by replacing the previous dynamic programming technique with graph cuts. Alternatively, an importance diffusion scheme is introduced by Cho *et al.* [22], aiming at emphasizing the pixels adjacent to the removed seams since the seams also contained context information. Although the existing seam carving based methods can have good performance in some cases, they may also produce seams cutting through important areas, which causes distortion and creates serious artifacts.

Though many methods have been developed to get an improved resizing result, the critical point remains unchanged, which is how to find the objects that should be maintained in the resizing process. Traditional attempts try to utilize low-level clues. However, these information has limited ability to guide the determination of important objects. Based on this consideration, saliency based clues [23], [24], [25], [26], which exhibit certain kind of high-level semantic meaning and indicate human attention, are favored recently. These clues can infer the probable salient regions and protect them from distortion. In this paper, the proposed method also employs the high-level
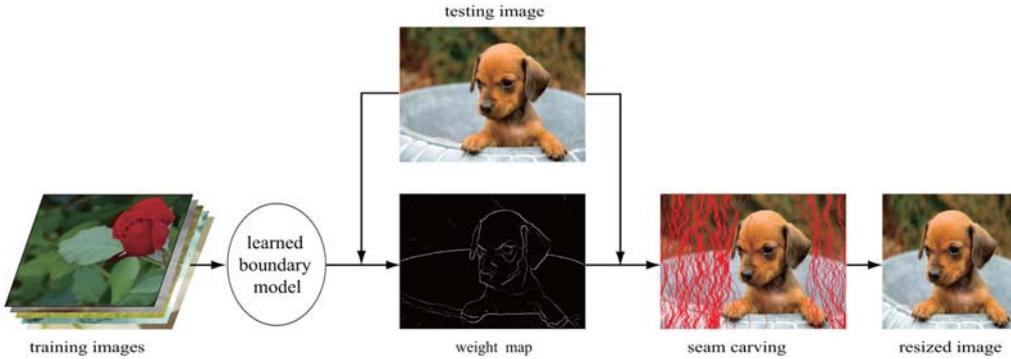
Figure 2: The flow chart of the proposed resizing method. In this process, a weight map, which is learned by a boundary model, is utilized as a prior information in the proposed framework.

clue, but in an alternative manner. We consider the semantic boundary of an image, which can capture the main structure of important objects instead. Detailed introduction is followed in the next section.

3. The proposed method

3.1. Overview

In this paper, a novel method based on machine learning is developed by incorporating high-level information of the ROI. It aims to overcome the distortion of the ROI in image resizing. Though the seam carving framework is taken as an example to illustrate the proposed idea, it can be straightforwardly extended to other resizing prototypes, such as warping and segmentation based ones.

The most popular work of *seam carving* by Avidan and Shamir [20] seeks the seams with global minimum energy in one direction via optimizing an energy function. Nevertheless, the obtained results are not satisfactory due to limited use of ROI priors. If the prior information of the interesting objects is utilized, the generated seams cutting through the ROI may decrease. Based on this consideration, the boundary information [27] of the ROI is employed as an guidance. Firstly, the boundary model is learned by the logistic regression. Then each input image is processed according to the obtained model to get the extracted boundary. This semantic clue is finally applied in the energy definition of seam carving, which indicates the existence of ROI. A flow chart is shown in Fig.2, and the concrete introduction is as follows.

Boundary learning. Boundary is different from traditionally known edge. An edge is most often defined as an abrupt change in some low-level image features such as brightness or color, while a boundary is a contour of high-level information that represents a change in pixel ownership from one object or surface to another. Given training images with labeled ground truth, a boundary model is learned by classical *logistic regression* with selected image features. The employed features are brightness, color, and texture.

Image resizing. For an input image, features are firstly extracted. Then the learned boundary model is employed to extract boundaries of interesting areas. After that, a weight map is calculated according to the obtained boundary and the seams are carved by an energy minimization procedure.

3.2. Boundary learning

This section introduces the soft classification model employed in the boundary learning process. Since judging whether a pixel is a part of a boundary is actually a classification problem, a classification model is required to learn a combinational rule of various image features from the training data. According to [28], logistic regression model can be more easily implemented and has faster convergence rate than *support vector machines*, *classification trees*, *density estimation*, and *hierarchical mixtures of experts* [29]. Consequently, logistic regression is adopted for its superiority in this paper.

In fact, logistic regression model is one of the most widely used classifiers. In statistics, it is a part of statistical models called generalized linear models [30], [31]. An explanation of logistic regression begins with the logistic function, which is defined as:

$$f(X) = \frac{e^X}{e^X + 1} = \frac{1}{1 + e^{-X}}, \tag{1}$$

where the variable X represents some set of independent variables, while $f(X)$ represents the probability of a particular output, ranging from 0 to 1. The input X sometimes is the linear combination of several independent variables. It is usually defined as:

$$X = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots \alpha_k x_k, \tag{2}$$

where x_i is the independent variable.

To learn these coefficients, the ground truth labeling of a pixel being on the boundary or not is required. Then the vector of coefficients $(\alpha_1, \cdots, \alpha_n)$

is learned by maximizing the likelihood with Newton-Raphson. In our application, the independent variables represent the distinct features extracted from images. When the relationship of those features is learned, it is used in the model to get a predicted value for every pixel in the test images. The predicted value ranges from 0 to 1, indicating the possibility of being a boundary. The higher the predicted value is, the more probable the pixel is in the boundary. In the following, we will explain the feature vector X employed in the model in detail.

3.3. Image features

With the logistic regression model discussed above, the boundary of an image can be extracted based on the image features. Brightness, color, and texture are the most effective ones, which are used in this paper to get a content-aware boundary. With these low-level features, a top-down prior can be obtained. The parameters for describing these features are chosen with respect to training data. After obtaining the image features, feature vectors are formed to train and test the classifier constructed by a logistic regression model.

3.3.1. Brightness and color

Brightness and color are the most commonly employed features for describing images. As for their representation, there are several choices of color spaces. Some researchers claim RGB is suitable because it is computationally efficient and well designed for image display. Other researchers claim HSV is better since it provides an intuitive representation of color and models the way in which humans perceive color. A number of researchers choose the $L^*a^*b^*$ space for the reason that it is quite similar to human perception. However, there are no experimental justification for their choices. In order to get the best color space in this paper, the above mentioned three spaces are all evaluated. Their results are shown in Fig.3, which demonstrates the performance of employing each color space individually for the seam carving task. From the curves in Fig.3, it is clear that $L^*a^*b^*$ color space performs best because it produces the minimum error. To be specific, for the $L^*a^*b^*$ color space, L^* corresponds to brightness, and a^* and b^* represent two color channels respectively. The adoption of $L^*a^*b^*$ space is motivated by the fact that the a^* and b^* channels correspond to the perceptually orthogonal red-green and yellow-blue color opponents found in the human visual system

(see Palmer [32]). In the proposed method, $L^*a^*b^*$ is selected, and each image is converted into $L^*a^*b^*$ color space to compute brightness and two color gradients.
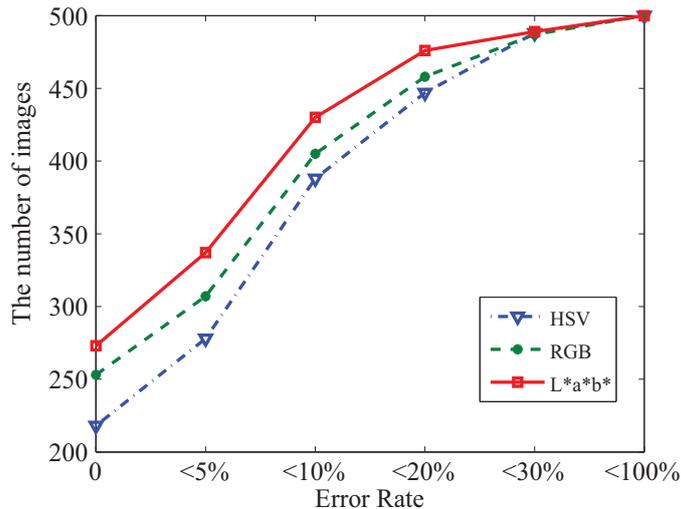

Figure 3:   A quantitative comparison of HSV, RGB, and $L^*a^*b^*$ color spaces. The $L^*a^*b^*$ achieves the lowest error rate in image resizing compared with the others.

3.3.2. Texture

Texture feature is another type of clue to express image content. Typically, texture extraction involves computationally intensive expression for each pixel [33]. One of the most widely used way to describe texture is to project the image intensity to a basis function set. Since each function in the basis function set is in a different frequency domain, the projection is referred to as spectral decomposition [34]. Then the image texture can be represented by those basis functions. The coefficients projected onto those basis functions correspond to the energy of the texture in a specific scale and orientation.

One typical form for the basis function set is a series of filters, which is often called a filter bank. Approaches in various literatures have been proposed to convolve images with a bank of linear spatial filters with respect to different orientations and spatial frequencies [35], [34], [36]. In this paper, the oriented filter bank is based on a Gaussian derivative and its Hilbert transform. The use of the Hilbert transform instead of a first derivative
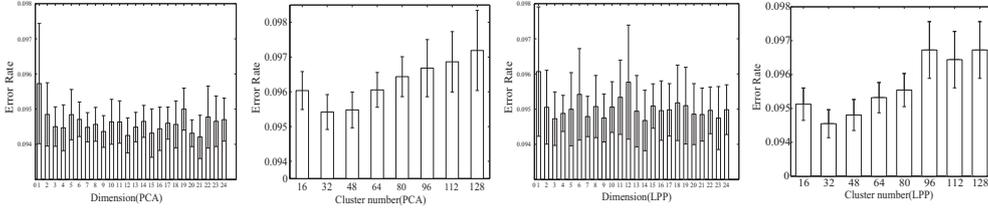
8

Figure 4: The statistical results of the error rates with different numbers of reduced dimensions and clusters. The first pair are the error rates with different dimensions and clusters for PCA. The best dimension and cluster number is 12 and 32, respectively. The second pair present the error rates for LPP. The best dimension and cluster number is 14 and 32, respectively.

makes f_1 and f_2 an exact quadrature pair. To be more precise, the expression is as follows:

$$
\begin{aligned}
f_1(x,y) &= \frac{d^2}{dy^2}(\frac{1}{C}exp(\frac{y^2}{\sigma^2})exp(\frac{x^2}{l^2\sigma^2})), \\
f_2(x,y) &= Hilbert(f_1(x,y)),
\end{aligned}
\tag{3}
$$

where σ stands for the scale of the filter, l is the aspect ratio of the filter, and C is a normalization constant. Each filter is zero-mean and L_1 normalized for the sake of scale invariance [37].

In order to extract texture, each image is converted into gray scale at first. Then the image is filtered by the filter bank to obtain a response vector for each pixel in the image. Although using filter responses to represent the textures is easy to implement, the representation is redundant since these are overlaps between the filters. This observation leads to the proposal of dimensionality reduction at first and then clustering the filter responses into several groups.

To this end, two strategies for dimensionality reduction are adopted: *principle component analysis* (PCA) and *locality preserving projection* (LPP). By reducing the dimensionality of the representation, we can decrease the computational complexity without losing the desirable information. In order to get the reasonable numbers of dimensionality and clusters, an initial filter bank of 24 filters with respect to specific scales and orientations are chosen. 500 images are then used to train the best parameter. Fig.4 presents the error rate (it will be defined in Section 4) under different clustering numbers and vector dimensionalities. In the first pair of Fig.4, the lowest error rate of PCA is at 12 dimensions and 32 groups. It can also be observed from

9

the second pair that the lowest error rate of LPP is at 14 dimensions and 32 groups. As a result, the best number of cluster is 32 in both Euclidean space and manifold space. However, the best dimensions for PCA and LPP are different. The experimental results also show that the error rates by the two different dimension reduction methods are similar. Consequently, the linear space is preferred in the proposed method.

3.4. The resizing methodology

The general resizing procedure employs gradients as the indication of ROI to be preserved. The involved energy definition is usually as follows

$$e_g(x, y) = |\frac{\partial}{\partial x}I(x, y)| + |\frac{\partial}{\partial y}I(x, y)|, \tag{4}$$

where $e_g(x, y)$ represents the energy value at (x, y), and I is the gray scale of the image. Greater value means larger possibility of being an important object. But this low-level clue is not capable enough to serve as guidance, which leads to distortions and serious artifacts. In the proposed method, I is the boundary map extracted from the original image. Assume that the image width needs to be reduced. The cost of removing a seam is defined as $E(s) = \sum_{i=1}^{n} e(s_i)$, where s_i is the ith pixel in a vertical seam, and n is the number of pixels in vertical direction. If a seam passes through the boundary, the energy function will become larger. The optimal seam s^* can be obtained by minimizing this energy function:

$$s^* = arg\min_{s} E(s) = arg\min_{s}\sum_{i=1}^{n} e(s_i). \tag{5}$$

4. Experiments

In this section, several experiments are conducted to evaluate the performance of the proposed method for image resizing. Firstly, the employed data sets are introduced. Then the evaluation metrics are established to get a quantitative measure. In the end, the experimental results are analyzed in detail.

Figure 5: Some sample images in the Data1. The first row are the original images and the second row are their ground truth.

4.1. Data set

In order to evaluate the performance of the proposed algorithm, two data sets are involved in the evaluation procedure. The first one was constructed by Achanta *et al.* [38] and has achieved great popularity in image processing community. It contains 1,000 images and every image in the data set has a ground truth delineation of ROI. In this case, the experimental results can be evaluated quantitatively. In our settings, 500 images are randomly chosen for training boundary model and the other 500 for testing the proposed method. Some sample images and their ground-truths are shown in Fig.5. The second data set was established by Michael *et al.* [39] and it is generally used for subjective user study. There are 80 images in together and no ground truth reference is provided. The two data sets are respectively denoted as *Data1* and *Data2*.

4.2. Evaluation measure

Since the two data sets are respectively used for objective and subjective evaluation, their corresponding metrics are not the same. As for the first data set, experimental results are compared with the reference ground truths. Although there is a quantitative evaluation [40], it is difficult to operate and hard to evaluate on a large data set. In order to get a reasonable and plausible measure, an error rate is defined in this paper as follows:

$$error\ rate = \frac{across\ seams}{total\ seams} \times 100\%, \tag{6}$$

11

where across seams stand for the number of seam pixels cutting through the ground truth salient objects, and total seams represent the total number of pixels in the removed or added seams. This index is useful for most of images, and can provide a fast and relatively precise evaluation.

As for the second data set, the experimental results are firstly judged by 15 participants, each of which gives every result a score to indicate the goodness of the resizing result. The score ranges from 1 to 5 with higher scores mean better resizing results. Then the consistence among the participants' evaluations are justified using one form of the *Intraclass Correlation Coefficient* psychological model [41], [42]. To be specific, the $ICC(3, k)$ is employed for the task. It measures the consistency of the k participants' mean ratings and is defined as

$$ICC(3, k) = \frac{bms - ems}{ems}, \tag{7}$$

where bms represents mean square of the ratings between targets, ems means total error mean square, and $k(= 15)$ is the number of participants. The values of ICC range from 0 to 1, where 0 implies no consistency and 1 complete consistency.

4.3. Results on Data1

In this part, experimental results on *Data1* are shown and compared with traditional seam carving method of Avidan and Shamir [20]. All the images are resized horizontally to 75% width of the original size. Fig.6 demonstrates some examples of the resized images generated by the two methods. It can be seen that in the proposed method, the seams which cut through the ROI, are much fewer than those produced by [20]. Compared with the traditional method, the proposed method can preserve the ROI better. Furthermore, the images obtained by the proposed method can better blend into the background and highlight the ROI. The reason for the success of the proposed method is mainly that the learned boundary can effectively indicate the ROI when removing seams. Though the traditional seam carving is also content-aware based method, it uses low-level image clues that can hardly reflect the ROI information. With a higher-level prior of the boundary, the proposed method is superior in maintaining the ROI compared to the traditional seam carving method.

In order to evaluate the method quantitatively, Fig.7 shows a statistical result of the proposed method compared to that of the traditional one.

Figure 6: Comparison between the original seam carving method [20] and the proposed method. (a) Original images; (b) Results of removed seams by [20]; (c) Results of resized images by [20]; (d) Detected boundaries from original images; (e) Results of removed seams by the proposed method; (f) Results of resized images by the proposed method.

The comparative results give a quantitative understanding of the proposed method. From the statistical results, the proposed methods with PCA and LPP based texture expressions are almost the same, both of which outperform the traditional one. Specifically, the number of the images whose error rates are less than 5% exceeds half of the testing images, which is 100 more than that of traditional seam carving method. Although different spaces have different distance measurements, their best solutions are all satisfying for the proposed method.
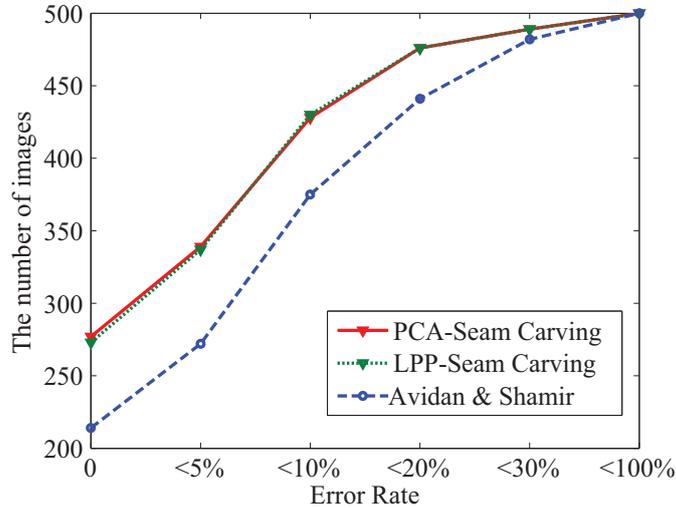

Figure 7:   A quantitative comparison of the proposed method and [20] on 500 testing images.

4.4. Results on Data2

In this part, subjective experimental results on *Data2* are shown to justify the effectiveness of the proposed method. Each image is firstly resized to the 75% height of the original size by the proposed method. Then the obtained result and the original image are displayed simultaneously to be evaluated by different participants. As for the criterion of assessment, no more principles are given except that the participants are expected to score the resizing resultss according to their visual experience. After that, the ratings from different participants are statistically analyzed according to the measure introduced in Section 4.2.

14

Fig. 8 illustrates several typical examples of the resized results. The averaged score from different participants is 4.0 for the proposed method, with a high consistency indication of $ICC(3, 15) = 0.94$. This means the participants' evaluation is highly consistent and the obtained results are convincing. We further compare the proposed method with traditional seam carving, whose averaged score is 3.6 and $ICC(3, 15) = 0.95$. This is obvious that the proposed method is better than the traditional seam carving.
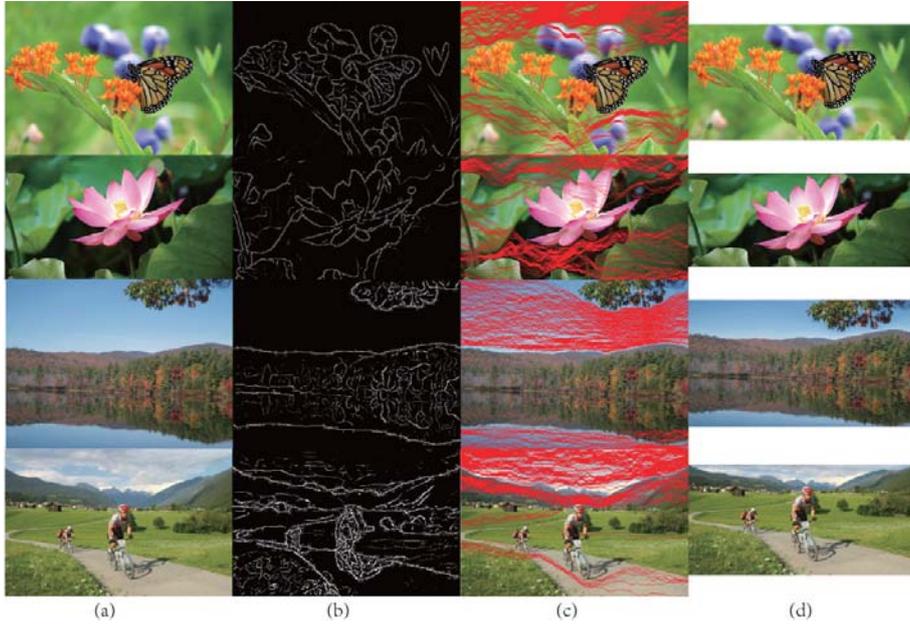

Figure 8: Illustrative results of the proposed method on Data2. (a)Input images. (b)Detected boundaries. (c)Removed seams. (d)Resized results.

5. Discussion

In this section, several issues related to the proposed method are discussed. They are respectively the robustness to noise, superiority compared with edge based and saliency based methods. In the end, several failure cases are shown to illustrate the limitation of our method.

5.1. Robustness to noise

To evaluate the robustness of the proposed method, different levels of Gaussian noise are added into one channel of the original color image. Then

15

the proposed method is applied to evaluate the resized results. The Gaussian noises are set as 20db, 40db, 60db, 80db and 100db, respectively. Fig.9 shows the resized results by the proposed method on the *fire balloon* image with different levels of Gaussian noise. Recall that the resizing results using PCA and LPP are almost the same. Thus we only present the results by using PCA. It can be easily found that there are only a few seams cutting through the ROI despite the influence of noise. The interesting regions in the resized images are all well preserved when the noise increases from 0db to 100db. The main reason for this phenomenon is that the prior information is utilized as a weight map, which is a global information of the ROI with little influence under noise. This can guide the seams to bypass the ROI.

Furthermore, in order to give an accurate evaluation of the proposed method, the quantitative results compared with the traditional seam carving method are also presented in Fig.10. Although the error rates increase as the noise increases, the proposed method performs better than the traditional seam carving. Specifically, when the noise level increases from 0db to 100db, the increase of error rates is less than 3% using either PCA or LPP. This further demonstrates the robustness and effectiveness of the proposed method.

5.2. Boundary versus edge

In this part, the proposed boundary based method and the classical Canny edge based method are compared as different means of energy definition. Note that Canny detection is not a learning based method. In fact, Canny edges can also be interpreted as a low-level feature which indicates the changes of pixel intensities. Fig.11 shows some typical results generated by the proposed method and Canny based method. Similarly with previous discussion, the proposed method only demonstrates PCA based results. It can be easily observed that the proposed method achieves better performance compared to the Canny based method. Much fewer seams cut through the ROI and the the main objects are generally well preserved. But some of the results produced by Canny based method are severely distorted, especially the *girl* image in the third row.

To further illustrate the effectiveness of the proposed method, statistical results of the two methods are presented in Fig.12. It can be found that both of the prior based methods can achieve better performance than the original seam carving method. It means that boundary or edge is a useful prior for image resizing. Though both the prior based methods are better

Figure 9: An example performance of the proposed method on noisy data. The added Gaussian noise in one color channel is (a)0db, (b)20db, (c)40db, (d)60db, (e)80db, and (f)100db. For each row, the first image is the original image, the second one is the removed seams and the last one is the resized image.
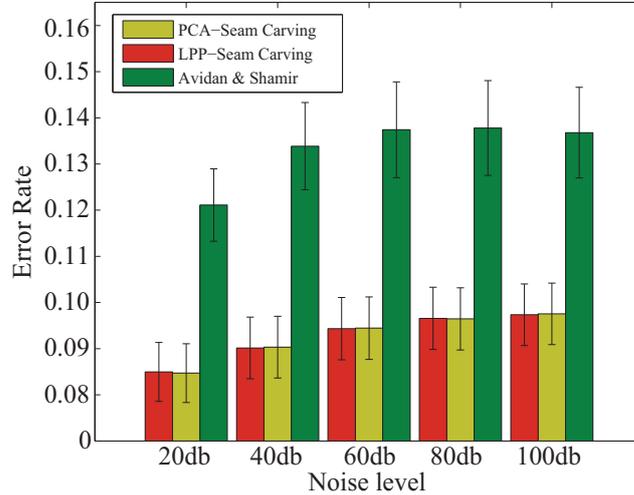
Figure 10: Comparison of error rates under different Gaussian noises. With the increase of noise level, the averaged error rate is increasing. But the increase is less than 3% for the proposed method on all the noise levels, which is more robust than the original seam carving method.

than the original one, the boundary based method is more promising than the Canny based one. For example, the number of the images, whose error rates are less than 5%, are 80 more than that of Canny based method in Fig.9. Since the Canny edge is not a high-level information, it will mislead the optimization of the energy function. On the contrary, the proposed method is more informative, which leads to a more satisfying resizing result.

5.3. Boundary versus saliency map

In the proposed method, the boundary of ROI is employed as the prior information to guide image resizing. In fact, there are other features of ROI, such as saliency map [47], which can also be used as prior information. Fig.14 gives some resized results by using several popular saliency detection methods AC[25], CA[26], FT[43], GB[44], IT[45] and SR[46]. It can be seen that the saliency map of ROI is also the useful prior information for the problem of image resizing. The better the saliency map for the ROI is, the better the resizing result is. At the same time, the results for the proposed method are better than other prior methods. Take Fig.13 for example. The flagstaff and the flower stem are one part of the ROI. But the resizing results
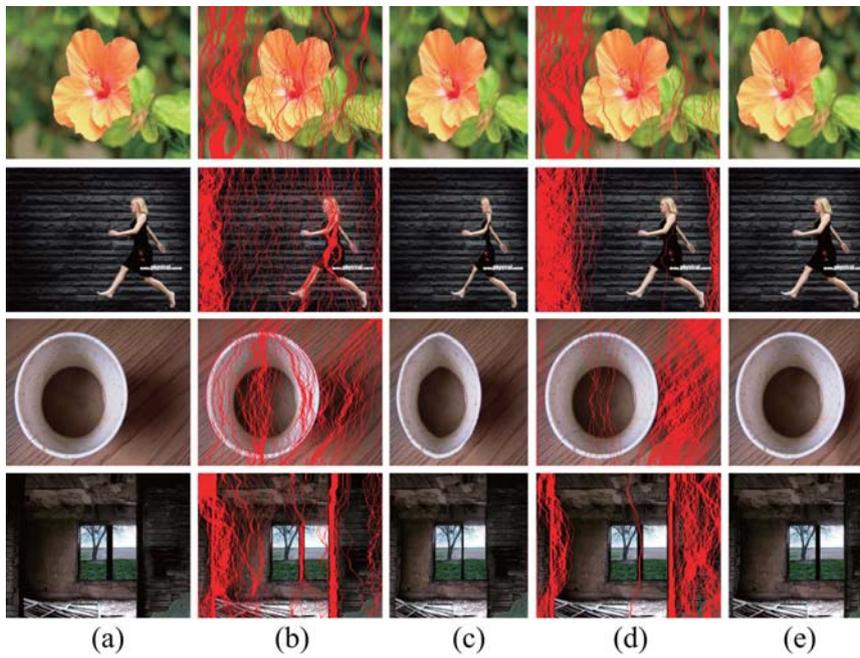
18

Figure 11: Comparisons between the proposed method and Canny based method. (a)the original images, (b)results of removed seams by Canny detection, (c)results of resized images by Canny detection, (d)results of removed seams by the proposed method, and (e)results of our resized images by the proposed method.
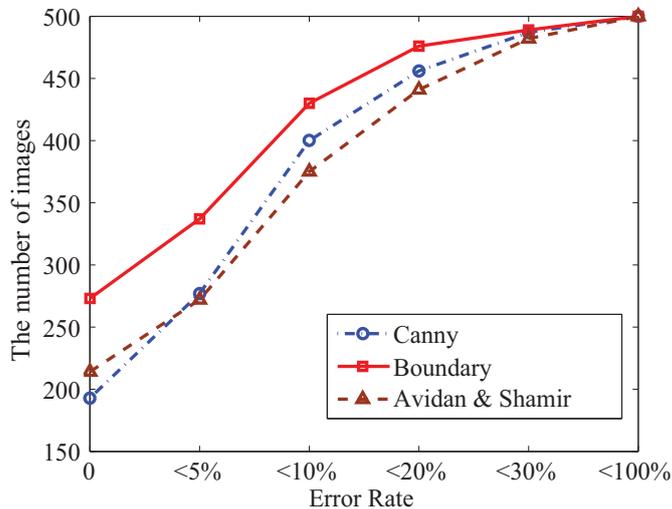
Figure 12: A quantitative comparison of the proposed method and and Canny based method on 500 testing images, respectively. The proposed method achieves lower error rate in image resizing compared with the Canny based one. The boundary detection is better than the Canny detection in performing seam carving.

by the proposed method are better than the saliency based resizing results. The saliency based resizing results are significantly distorted for the flagstaff and flower stem. The possible reason is that the boundary model adopted in the proposed method is a supervised learning procedure, which not only takes full advantage of the test image but also takes into consideration the information of the training data. While the aforementioned saliency models are unsupervised, which only consider the test image itself and mainly utilize the color contrast. Hence the boundary based model can provide a better prior information for image resizing.

5.4. Failure cases

The above experiments demonstrate the robustness and effectiveness of the proposed method. However, there are situations that it does not work satisfactorily. Fig.14 shows several typical failure cases of the proposed method. The left image illustrates an unclear dominate object. In this image, both the pins and the fastener tangle together and it is hard to tell which one is the most salient one. The middle image is a large plate, which almost fills the whole image. No matter what direction this image is to be resized, it
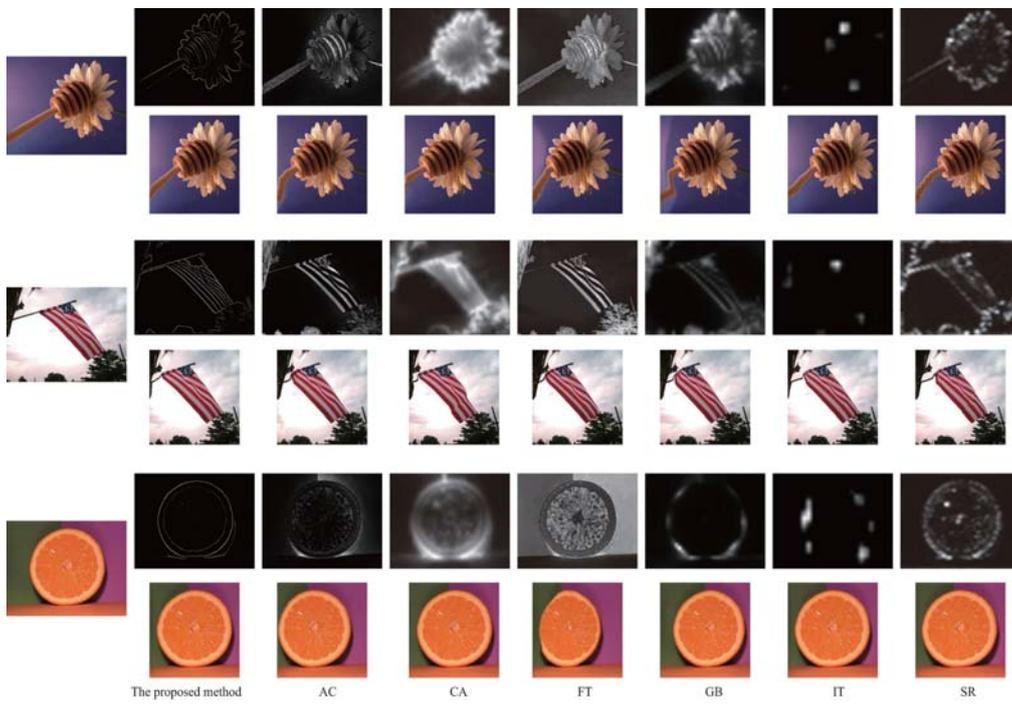
Figure 13: Comparison of image resizing results by using boundary and various saliency maps as the prior information. From left to right, each column respectively represents the original images, results of the proposed method, results by using saliency maps of AC[25], CA[26], FT[43], GB[44], IT[45], and SR[46].

Figure 14: Several typical example images on which the proposed method will fail.

will lead to distortion of the plate. The right one is an orange image, which is occluded by leaves. Since the boundary of the orange is incomplete, the orange is distorted in the process of resizing.

6. Conclusion

In this paper, a novel method based on supervised learning has been proposed for image resizing. The boundary model learns a rule of combining image features (brightness, color, and texture) by a logistic regression algorithm. Under this model, the learned boundary acts as the prior information to guide the process of resizing. Therefore, the important regions and the structural consistency of the input image can be preserved when the image is resized into the target size.

Though we formulate the proposed method in the context of seam carving, it does not means that the method is only limited to this case or is an improvement of seam carving. Our method can be readily extended to other resizing framework. For example, the warping based methods make the unimportant regions absorb most distortion and maintain the ROI as much as possible. In this situation, the learned boundary can serve as an implication of the ROI. As for the segmentation based or cropping based methods, the extension is similar because they all need a precise indication of ROI and the learned boundary can fulfill this task. However, we admit that there might be other high-level clues that can be incorporated into the resizing procedure. Our method is only an example of this attempt. Future work will investigate other high-level clues on this topic.

Acknowledgment

References

1. Vaquero, D., Turk, M., Pulli, K., Tico, M., Gelfand, N.. A survey of image retargeting techniques. In: *Proc. SPIE 7798, Applications of Digital Image Processing XXXIII*. SPIE; 2010, .

2. Ren, T., Liu, Y., Wu, G.. Image retargeting based on global energy optimization. In: *IEEE International Conference on Multimedia and Expo*. 2009, p. 406–409.

3. Wang, S., Lai, S.. Fast structure-preserving image retargeting. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009, p. 1049–1052.

4. Maybank, S.. A probabilistic definition of salient regions for image matching. *Neurocomputing* 2013;**120**:4–14.

5. Ban, S.W., Jang, Y.M., Lee, M.. Affective saliency map considering psychological distance. *Neurocomputing* 2011;**74**(11):1916–1925.

6. Temporal spectral residual for fast salient motion detection. *Neurocomputing* 2012;**86**:24–32.

7. Wang, Q., Zhu, G., Yuan, Y.. Multi-spectral dataset and its application in saliency detection. *Computer Vision and Image Understanding* 2013;**117**(12):1748–1754.

8. Chen, L., Xie, X., Fan, X., Ma, W., Zhang, H., Zhou, H.. A visual attention model for adapting images on small displays. *ACM Transactions on Multimedia Systems* 2003;**9**:353–364.

9. Ma, M., Guo, K.. Automatic image cropping for mobile device with built-in camera. In: *IEEE Consumer Communications and Networking Conference*. 2004, p. 710–711.

10. Suh, B., Ling, H., Bederson, B., Jacobs, W.. Automatic thumbnail cropping and its effectiveness. ACM; 2003, p. 95–104.

11. Amrutha, S., Shylaja, S., Natarajan, S., Balasubramanya, N.. A smart automatic thumbnail cropping based on attention driven regoins of interest extraction. In: *International Conference on Interaction Scaences*. ACM; 2009, p. 957–962.

12. Ciocca, G., Cusano, C., Gasparini, F., Schettini, R.. Self-adaptive image cropping for small diaplays. *IEEE Transactions on Consumer Electronics* 2007;**53**(4):1622–1627.

13. Setlur, V., Takagi, S., Rasker, R., Gleicher, M., Gooch, B.. Automic image retargeting. In: *International Conference on Mobile and Ubiquitous Multimedia*. ACM; 2005, p. 59–68.

14. Liu, H., Jiang, S., Huang, Q., Xu, C., Gao, W.. Region-based visual attention analysis with its application in image browsing on small displays. In: *International Conference on Multimedia*. ACM; 2007, p. 305–308.

15. Hasan, M., Kim, C.. An automatic image browsing technique for small display users. In: *International Conference on Advanced Communication Technology*. 2009, p. 2044–2049.

16. Liu, F., Gleicher, M.. Automatic image retargeting with fisheye-view warping. In: *ACM symposium on User interface software and technology*. 2005, p. 153–162.

17. Wang, Y., Tai, C., Sorkine, O., Lee, T.. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics* 2008;**27**(5).

18. Zhang, G., Cheng, M., Hu, M., Martin, R.. A shape-preserving approach to image resizing. *Computer Graphics Forum* 2010;**28**(7):1897–1906.

19. Jin, Y., Liu, L., Wu, Q.. Nonhomogeneous scaling optimization for realtime image resizing. *The Visual Computer* 2010;**26**(6-8):769–778.

20. Avidan, S., Shamir, A.. Seam carving for content-aware image resizing. *ACM Transactions on Graphics* 2007;**26**(3).

21. Rubinstein, M., Shamir, A., Avidan, S.. Improved seam carving for video retargeting. *ACM Transactions on Graphics* 2008;**27**(3):1–9.

22. Cho, S., Choi, H., Matsushita, Y., Lee, S.. Image retargeting using importance diffusion. In: *International Conference on Image Processing.* 2009, p. 977–980.

23. Wang, Q., Yuan, Y., Yan, P., Li, X.. Saliency detection by multiple-instance learning. *IEEE Transactions on Cybernetics* 2013;**43**(2):660–672.

24. Wang, Q., Yuan, Y., Yan, P.. Visual saliency by selective contrast. *IEEE Transactions on Circuits and Systems for Video Technology* 2013; **23**(7):1150–1155.

25. Achanta, R., Estrada, F.J., Wils, P., Süsstrunk, S.. Salient region detection and segmentation. *Computer Vision Systems* 2008;:66–75.

26. Goferman, S., Zelnik-Manor, L., Tal, A.. Context-aware saliency detection. In: *IEEE Conference on Computer Vision and Pattern recognition.* 2010, p. 2376–2383.

27. Kokkinos, I., Deriche, R., Faugeras, O., Maragos, P.. Computational analysis and learning for a biologically motivated model of boundary detection. *Neurocomputing* 2008;**71**(10C12):1798–1812.

28. Martin, D., Fowlkes, C., Malik, J.. Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions Pattern Analysis and Machine Intelligence* 2004;**26**(5):530–549.

29. Jordan, M.I., Jacobs, R.A.. Hierarchical mixtures of experts and the em algorithm. *Neural Computation* 1994;**6**:214–1994.

30. Friedman, J., Hastie, T., R.Tibshirani, . Additive logistic regression: a statistical view of boosting. *Annals of Statistics* 2000;**28**:377–386.

31. Schapire, R.E., Singer, Y.. Improved boosting algorithms using confidence-rated predictions. *Machine Learning* 1999;**37**(3):297–336.

32. Palmer, S.. *Vision Science.* Massachusetts Institute of Technology Press; 1999.

33. Zhang, H., Fritts, J.E., Goldman, S.A.. A fast texture feature extraction method for region-based image segmentation. In: *16th Annual*

Symposium on Image and Video Communication and Processing; vol. 5685. 2005, .

34. Malik, J., Belongie, S., Leung, T., Shi, J.. Contour and texture analysis for image segmentation. *International Journal of Computer Vision* 2001;**43**(1):7–27.

35. Srinivasu, P., Avadhani, P.S., Satapathy, S.C., Pradeep, T.. A modified kolmogorov-smirnov correlation based filter algorithm for feature selection. *Advances in Intelligent and Soft Computing* 2012;**132**:819–826.

36. Ahonen, T., Pietikainen, M.. Image description using joint distribution of filterbank responses. *Pattern Recognition Letters* 2009;**30**(4):368–376.

37. Perona, P., Malik, J.. Detecting and localizing edges composed of steps, peaks and roofs. In: *International Conference on Computer Vision*. 1990, p. 52–57.

38. Achanta, R., Süsstrunk, S.. Saliency detection for content-aware image resizing. In: *International Conference on Image Processing*. 2009, p. 1005–1008.

39. Rubinstein, M., Gutierrez, D., Sorkine, O., Shamir, A.. A comparative study of image retargeting. *ACM Transactions on Graphics (Proc SIGGRAPH Asia)* 2010;**29**(5):160:1–160:10.

40. Ren, T., Wu, G.. Automatic image retargeting evaluation based on user perception. In: *International Conference on Image Processing*. 2010, p. 1569–1572.

41. Shrout, P.E., Fleiss, J.L.. Intraclass correlation: Uses in assessing rater reliability. *Psychology Bulletin* 1979;**86**(2):420–428.

42. Wang, Q., Wang, Z.. A subjective method for image segmentation evaluation. In: *ACCV (3)*. Springer; 2009, p. 53–64.

43. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.. Frequency-tuned salient region detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2009, p. 1597–1604.

44. Han, J., Ngan, K.N., Li, M., Zhang, H.. Unsupervised extraction of visual attention objects in color images. *IEEE Transactions on Circuits and Systems for Video Technology* 2006;**16**(1):141–145.

45. Itti, L., Koch, C., Niebur, E.. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1998;**21**(11):1254–1259.

46. Hou, X., Zhang, L.. Saliency detection: A spectral residual approach. In: *IEEE Conference on Computer Vision and Pattern Recognition.* 2007, p. 2044–2049.

47. Wang, Q., Yan, P., Yuan, Y., Li, X.. Multi-spectral saliency detection. *Pattern Recognition Letters* 2013;**34**(1):34–41.

Qi Wang received the B.E. degree in automation and Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2005 and 2010 respectively. He is currently an associate professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His research interests include computer vision and pattern recognition.

Yuan Yuan is a full professor with the Chinese Academy of Sciences (CAS), China. Her major research interests include Visual Information Processing and Image/Video Content Analysis. She has published over a hundred papers, including about 70 in reputable journals, like IEEE transactions and Pattern Recognition, as well as conferences papers in CVPR, BMVC, ICIP, ICASSP, etc.