Strip Based Media Retargeting via Combing Multi-Operators

Hao Wu, Jia Wang, and Hongkai Xiong, *Senior Member, IEEE*

***Abstract*—Based on the recently prioritized objectives and the statistical findings, we propose a novel media retargeting approach which considers content loss and deformation in a visual friendly manner. First, we divide an input image into several strips according to the average strip energy density and the line structure information. Then, we find the maximum resizing lengths which makes the deformation acceptable and choose the proper retargeting operator for each strip according to the energy information. The process above will be repeated iteratively until a cropping window can be selected which makes the content loss acceptable. The experimental results show that our proposed method can protect more important content information than traditional cropping and make less deformation than previous content-aware media retargeting method.**

***Index Terms*—Media retargeting, strip process, multi-operator.**

I. Introduction

In recent years, content-aware media retargeting, which means resizing input media to aspect ratios and different display sizes while protecting the visual information, has became an important problem. Most of the previous media retargeting methods can be classified as discrete or continuous.

The discrete methods, which try to identify important regions of the media by discard pixels from the other regions, are usually implemented by cropping and carving. The cropping algorithm is used to find the optimal position of a cropping window based on specified energy function. The cropping method does not create any artifact of the media, the removal of content is the most challenging problem of this method. Avidan and Shamir proposed the seam carving method in [1], which removes or inserts seams repeatedly from the source image to reach the target size. Each seam changes the size of an image by one. Rubinstein *et al.* improves the seam carving method by introducing forward energy functions [2] for better retargeting results. They also generalized the seam carving to video retargeting by employing two-dimensional seam manifold. To optimize an energy defined directly between the input and output images rather than forward energy in [2], Mansfield *et al.* define an energy over a visibility map that can be optimized using seam carving operations in [3]. They define meida retargeting as a binary graph labelling problem and solve it via seam carving optimization by given energy terms from a well defined general family. Though seam carving is a simple and efficient operator for media retargeting, it has a number of drawbacks. Firstly, the preprocessing time for one image is large as it will use dynamic program many times. Secondly, seam carving and related pixel removal methods fundamentally cannot retarget smooth curves to smooth curves. This may cause dominated structure (lines and edges) distortion and the deformation of objects in source media.

Continuous methods, on the other hand, retarget an image by rearranging and merging the pixels in the source image to fit into the target size. Wolf *et al.*[5] describe a real-time video resizing system based on spatially varying warping by solving a sparse linear systems of equations. In their retargeting scheme, a pixel with high importance is mapped to a distance of approximately one from its left neighbor, while a pixel with less saliency is mapped closer to its neighbor. Wang *et al.* presented a method which allows important regions to scale uniformly and homogeneous regions to be distorted by mesh grid warping for image retargeting in [6] and spread it to a motion-aware video resizing scheme in [7]. Krähenbühl *et al.* in [4] describe a retargeting system that compute bilateral temporal coherence energy accordingly for warp computation and then use temporal filtering of the per-frame saliency map to account for the camera and scene motion. Though the warping based methods in general show better structure-preserving property than the carving methods and better content-preserving property than the cropping methods, the distortion of proportions in the source media and the deformation can be a big problem. Moreover, these methods can be complex to implement and usually for each target size, the optimization should be re-run from the beginning. The result will even fail to be similar as uniform scaling in some cases.

As protecting contents often have conflicts with inserting deformation, most state-of-the-art media retargeting techniques try to show as much content as possible from the original media by modifying the shape of the objects in the media as discussed above. However, a detailed comparison given in [8] has indicated that in most cases viewers prefer sacrificing content over inserting deformation to the media.

To solve this problem, in this paper, we propose a novel media retargeting scheme which inserts less deformation and distortion than most previous content-aware retargeting operators while protecting more important content than manual cropping. The proposed scheme divides the source media into several strips. For each strip, proper retargeting operator would be chose with some constraints to insert low deformation. This process can be repeated iteratively to balance the content

The authors are with the Department of Electronic Engineering of Shanghai Jiao Tong University, and the Key Laboratory of Digital Media Processing and Transmission, Shanghai, 200240, China (e-mail: {hanchuanderen, jiawang, xionghongkai}@ sjtu.edu.cn). The authors were supported in part by the NSFC under Grant 61271221, U1201255, and by the National Key Technology R&D Program of China (2013BAH53F04).

loss and deformation. After that, automatic cropping is used to resize the image to the target size. Technique in [9] also uses the idea of strip process in media retargeting. But it just uses uniformly scaling for resizing while the proposed scheme combining multi-operators to achieve better result as no single retargeting operator performs well on all media and all target sizes. Moreover, method in [9] just considers the content while the proposed scheme making a balance between content loss and deformation. The structure information can be also protected better in the proposed scheme.

The rest of this paper is organized as follows. Section 2 shows the details of our proposed retargeting algorithm for image. Section 3 extends this algorithm to video retargeting. Section 4 compares our approach to the current state-of-the-art techniques. Section 5 concludes the paper.

II. Proposed algorithm for image retargeting

Assume the size of the source image I is $W_s \times H_s$, the size of the target image is $W_t \times H_t$. For the sake of simplicity, we will just discuss the width reduction of the source image in the following, so $W_t < W_s$ and $H_t = H_s$. The reduction of the height and the enlarging can be done analogously. In the following, we describe the proposed image retargeting algorithm.

Figure 1 shows the flow chart of the proposed algorithm. First, the energy function of the source image is computed. Then, the proposed scheme partitions the source image into several strips. Proper retargeting operator in each strip is chosen respectively to ensure the deformation acceptable. This process is repeated iteratively to balance the content loss and deformation. Finally, to resize the image to the target size, automatic cropping is used.

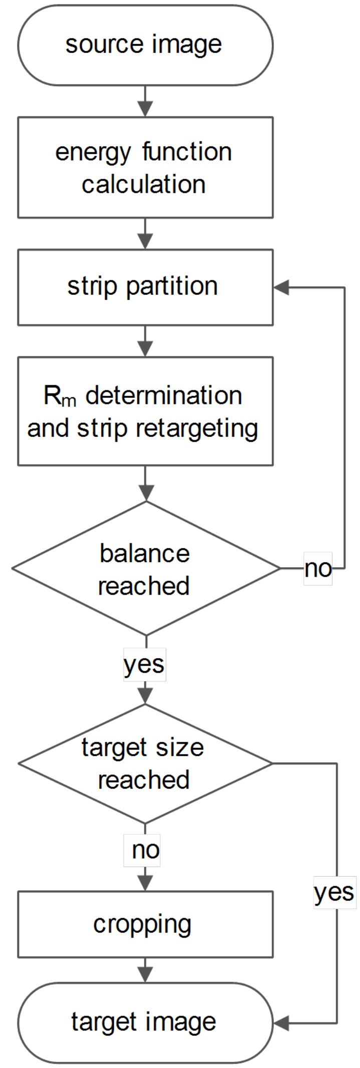


Fig. 1. The flow chart of the proposed algorithm. The balance between content loss and deformation decision procedure can make the proposed algorithm more visual friendly.

A. *Energy function*

To get the importance information of the source image, the energy of each pixel should be computed first. Though in most of the previous retargeting methods, gradient magnitude was chosen as the energy function, it can not protect enough important information when considering human attention model. Detecting salient areas of the images at object level should be valuable in the retargeting scheme, and as the deformation of faces is one of the most disturbing problem according to the survey in [8], two techniques, namely saliency map [10] and face detection [11], are employed in the proposed algorithm to generate the human attention information. So the energy function can be modified to:

$$E(x, y) = w_g \cdot Grad(x, y) + w_s \cdot sal(x, y) + w_f \cdot face(x, y) \quad (1)$$

where w_g ,w_s ,w_f are the weight of the gradient magnitude, saliency map and face map respectively; $Grad(x, y)$ and $Sal(x, y)$ are the gradient magnitude map and saliency map ranged from 0 to 1; $face(x, y)$ is the face map, which takes value 1 for the pixel belongs to a face and 0 otherwise.

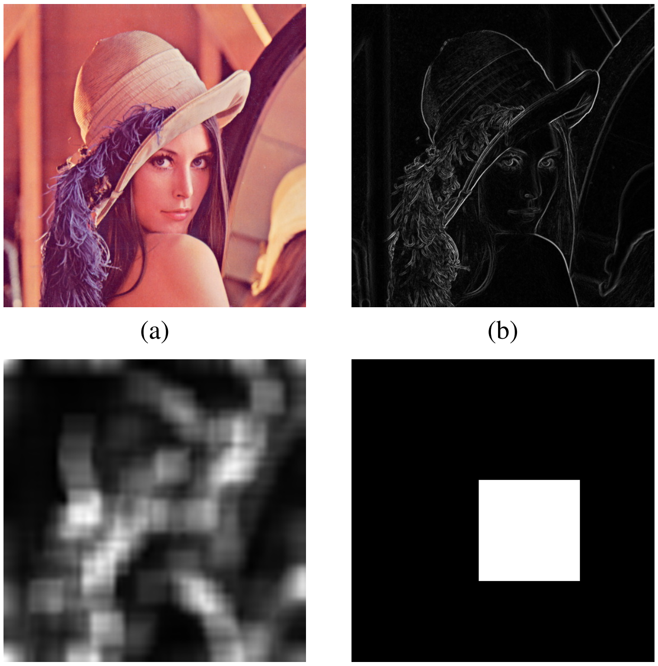
Figure 2 is an example of the energy function of the lena image.

As for the computational cost, it should be noted that the computational cost of the saliency map achieved in [10] is high. To decrease the computational complexity, we take the saliency detectors constructed from the phase spectrum of the image's Fourier transform [15] in practice. This approach is simple to implement and has nearly the similar result as [10]. The other maps (gradient and face) have very low computational complexity. Thus, the energy map achieving step can be implemented in real time.

B. *Strip Partition*

After finding the energy function, the complexity of the x-th column $e(x)$ can be computed by summing up the energy of pixels in the same column. Assume the source image is divided into N strips, the i-th strip ($0 \le i \le N-1$) has the left boundary l_i and right boundary f_i. We define the average energy density of the strip is the average complexity of all the columns in the corresponding strip. Then the optimal partition can be achieved by using iterative method.

Initially, the image is divided into strips of the same size. In the iterative procedure, assume the left boundary of the



(a) (b)

(c) (d)

Fig. 2. An example of the energy function. (a) The original image. (b) The gradient map. (c) The saliency map. (d) The face map. The detected face is marked as a rectangular region with high energy.

i-th strip l_i and the right boundary of the $(i+1)$-th strip f_{i+1} are fixed, then the proposed method will update the right boundary of the i-th strip f_i to get the new partition. The optimal partitioning should make the adjacent strips have average energy density as different as possible while the complexities of columns in the same strip as similar as possible. Based on these two criteria, one strategy is to minimize the variance of the column complexity in each strip while maximize the difference of the average energy density between two adjacent strips. The iterative updating boundary procedure can be formulated as:

$$\begin{aligned} f_i = \underset{l_i<f<f_{i+1}}{\arg\min} \, (& \sum_{x=l_i}^{f-1}(e(x)-avg_i)^2/(f-l_i) \\ & + \sum_{x=f}^{f_{i+1}-1}(e(x)-avg_{i+1})^2/(f_{i+1}-f) \\ & + \alpha \cdot \frac{1}{(avg_{i+1}-avg_i)^2}) \end{aligned} \tag{2}$$

where avg_i is the average energy density of the strip ranged from l_i to f, avg_{i+1} is the average energy density of the strip ranged from f to f_{i+1}, and α is a user specified weighting coefficient. In this work, α is fixed to 0.1 and the minimum length of the strip will be set to 10 pixels in practice, empirically.

After f_i is updated, the new f_{i+1} can also be updated with fixed f_i and f_{i+2} in a similar manner. This process is iteratively applied until all boundaries converge.

It should be noted that the partition above may cause severe line structure distortion. To solve this problem, straight line detection has been included in the partition scheme. Canny edge detector and Hough transform are used to identify straight lines. After finding the optimal partition discussed above, several adjacent strips which contain the straight lines will be combined together with some constraints.

C. Strip Retargeting

After dividing the source image into several strips, the proposed algorithm will retarget each strip individually by choosing seam carving or uniformly scaling operator. To insert less deformation into the source image the proposed algorithm will not resize the image to width W_t directly. The maximum removed width in this step is R_m, namely, $R_m \le W_s - W_t$. The remaining columns will be removed by cropping operator.

Now assume R_m has been given (The determination of R_m will be discussed in the next subsection), the strip retargeting problem can be equalized as finding the numbers of reduced columns in each strip. Assume that the i-th strip should reduce r_i columns in this step, then it can be computed as:

$$r_i = \lambda r_{i1} + (1-\lambda) r_{i2} \tag{3}$$

where λ is a user specified parameter ranged from 0 to 1.

In (3), r_{i1} and r_{i2} are number of strip reduced columns based on different strategy. For content aware retargeting, the strip with less average energy density will remove more columns. Assume the length of the i-th strip is len_i and the average energy density of this strip is avg_i. In [9] J.-S. Kim *et al.* have proposed one strategy to minimize the distortion in frequency domain via Fourier analysis based method, and formulate the number of strip reduced columns as:

$$r_{i1} = \max\left\{0, \min\left\{len_i, len_i(1+\frac{avg_i}{2\pi}\ln \mu len_i)\right\}\right\} \tag{4}$$

where $\mu \in (0, 1/\max_i len_i]$ can be achieved by the bisection search method according to the strip length information.

Though the distortion free strategy (r_{i1}) above can formulate the content distortions effectively, it may fail when considering structure distortions especially line structure distortions. To protect the line structure information in a content-aware way, the proposed algorithm adds anthoer strategy, the line protecting strategy, which can be formulated as:

$$r_{i2} = \max\left\{0, \min\left\{len_i, R_m/(avg_i \sum_{i=0}^{N-1}(1/avg_i)\right\}\right\} \tag{5}$$

Then r_i is a linear combination of r_{i1} and r_{i2} for the purpose of less distortion and better line structure information. Figure 3 gives an example of using line protecting strategy in the retargeting scheme to make more visual friendly result.

After finding r_i, the proposed algorithm will choose a proper operator for each strip based on some constraints. For the sake of simplicity, we use a width change threshold to achieve this goal as seam carving is better suited for small amount of changes. Strips with small width change (below the width change threshold) are resized by seam carving, otherwise uniformly scaling. For the purpose of real time ability, the proposed scheme uses real time seam carving algorithm in [12] rather than traditional seam carving in

(a) (b) (c)

Fig. 3. A comparison result about line structure information. (a) The source image. (b) and (c) are both the target image resized to 75% width with and without the line protecting strategy, respectively. We can see, result in (b) with the line protecting strategy is more visual friendly.

practice. As the computational cost of both seam carving and uniform scaling operator in this step is very low, the real time ability in this step can be guaranteed.

D. Balance between content loss and deformation

As the proposed algorithm is designed with the preference of content loss over the insertion of deformation, how to balance these two factors is the most challenging problem. To solve this problem, we use R_m to control the deformation in the proposed algorithm.

Defining the deformation factor of the i-th strip d_i is the rate between r_i and len_i, namely:

$$d_i = r_i / len_i \tag{6}$$

The strips with average energy density which exceeds the middle value of all the strips are defined as obvious strips. d_{thr} is a dynamic threshold which can be limit the deformation in the proposed algorithm:

$$d_{thr} = (1 - W_t / W_s)/2 \tag{7}$$

To make the deformation acceptable, R_m can be achieved in an iterative way by limiting the deformation factor of these obvious strips under the dynamic threshold d_thr defined above based on the change of the target size. The detail procedure to achieve R_m can be seen in algorithm 1.After finding R_m, strip retargeting procedure can be applied according to the discussion above.

Algorithm 1 Algorithm to achieve R_m

```
Initialize R_m = W_s − W_t, flag = 0; Compute r_i and d_i
  according to (3) and (6)
while flag = 0 do
  for each i ∈ [0, N − 1] do
    if the i-th strip is the obvious strip then
      if d_i > d_thr then
        This strip may insert deformation. So the value
        R_m should be decreased as R_m = 0.9 ∗ R_m,
        and change the flag value to 2.
      end if
    end if
    if flag = 2 then
      Some obvious strips may insert deformation,
      change back the flag value to 0 to continue the
      loop.
    else
      All obvious strips would not insert deformation.
      Then the flag value will be set to 1 to end the
      iteration.
    end if
  end for
end while
```

Then cropping operator is used to achieve the target size. But applying cropping operator directly may cause severe content loss. To solve this problem, the proposed algorithm will determine whether the strip retargeting procedure should be repeated before cropping. Considering the columns outside the potential cropping window, we define the column as important column if this column is in the strip with a specified high average energy density. The percentage of the important column is used to balance the content loss and deformation. If the percentage of important columns exceeds some threshold, the content loss can be determined as unacceptable and the strip retargeting procedure discussed above will be performed repeatedly until the percentage bellows this threshold before cropping.

The proposed algorithm can achieve better visual friendly results than most state-of-the-art media retargeting methods as it makes better balance between content-loss and deformation. Moreover, the computational complexity of the proposed algorithm is very low. In fact, it can retarget images to aspect ratios in real time.

III. Extension to video retargeting

In this section, we extend the proposed algorithm to video retargeting. We assume that the size of the source video is $Ws \times H_s$, the number of the video frames is T, and the size of the target image is $W_t \times H_t$. Again, for the sake of simplicity, we only discuss the reduction of frame width.

The video energy function is similar to 1, but the frame saliency map will be computed in a spatial-temporal way for the purpose of temporal smooth. In the proposed video retargeting scheme, the video saliency map is achieved by using the phase spectrum of quaternion Fourier Transform (QPFT) method [13].

After determining the energy function, video retargeting operator is applied. Similar to the image retargeting case, the retargeting process include strip retargeting and cropping optimization. To improve the jitter artifacts, we just apply the same partitioning and the same retargeting operator for the same strip to all frames in a video sequence in strip partition and strip retargeting process. For the sake of temporal coherence and real time ability, only uniformly scaling operator would be chose in strip retargeting procedure. As the proposed algorithm limits the deformation factor of the obvious strips to a narrow range, the scaling factor of all obvious strips in all frames would be similar. Fast moving objects across adjacent strips in adjacent frames can be retargeted effectively. This process is repeated until the cropping optimization can cause acceptable content loss which is similar to the image retargeting case.

To protect the temporal coherence and content aware, the cropping optimization will be solved as finding the best cropping panes path by considering the whole video content. The best path of the cropping panes should have the minimum sum of the energy removed outside the cropping panes of all path positions.

Assume W_{x_i} is one of the candidate cropping window ranged from x_i-th column to $(x_i + W_t - 1)$-th column in the i-th frame ($1 \le x_i \le W_s - W_t + 1$), which can be seen in Figure 4. $R^i_{x_i}$ is the total energy removed outside the window W_{x_i}, $\hat{R}^i_{x_i}$ is the minimum sum of the energy removed outside the cropping panes from the 1st frame to the i-th frame while having cropping window W_{x_i} in the i-th frame.

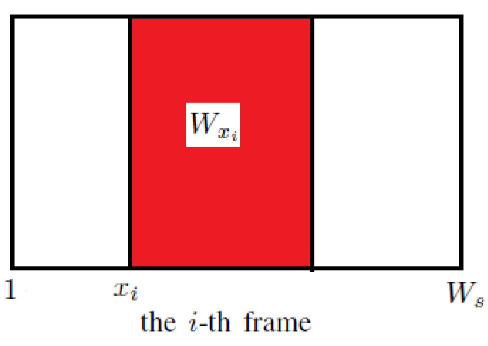


Fig. 4. Example of the candidate cropping window started from x_i-th column in the i-th frame, the whole region is the i-th frame with width W_s , the red region is the candidate cropping window. Different x_i can result in different cropping windows.

As discussed above, finding the best path of cropping panes can be equalized as finding the minimum sum energy outside the cropping panes over all frames, namely, finding $\hat{R}^i_{x_i}$. This can be solved via dynamic programming. The objective function can be formulated as:

$$\hat{R}^i_{x_i} = R^i_{x_i} + \arg\min_{x_{i-1}} \{\hat{R}^{i-1}_{x_{i-1}} + \beta \, |x_{i-1} - x_i|\} \tag{8}$$

where $\hat{R}^{i-1}_{x_{i-1}}$ is the minimum sum of the removed energy from the 1st frame to the $(i-1)$-th frame, x_{i-1} is the started position of the cropping window in the $(i-1)$-th frame. $|x_{i-1} - x_i|$ is the temporal smooth penalty function, and β is a user specified parameter to balance spatial content and temporal coherence. For the sake of simplicity, distance of the cropping position between adjacent frames can be set as less than 10 pixels, namely, $|x_{i-1} - x_i| < 10$ in our experiment. The cropping pane in the 1st frame is automatically chosen similar to image retargeting to achieve $\hat{R}^1_{x_1}$ for initialization. By backtracking from this minimum, we find the optimal path for a cropping panes.

The resulting cropping panes path is nearly temporally coherent as the distance of the cropping pane between two adjacent frames is very small, but it may still contain jitter. In order to lessen the effect of jitter, a Gaussian filter can be used to smooth the cropping position of each frame in the best cropping panes.

IV. Experimental results

We test our proposed method on the benchmark data set provided by Rubinstein *et al.* [8]. We compare our results to 3 best techniques, namely, multi-operator retargeting in [14], manually cropping and streaming video in [4], as is suggested in [8]. A no-reference comparison where the original image was not shown is implemented. 5 volunteers are invited for the comparison and the result is ranked from 1 to 4 for each operator. 1 means the result is the best while 4 the worst. Each volunteer will compare 12 image sets randomly chosen out the data set. The comparison result is shown in Table I.

TABLE I
Comparison result of four evaluated operators.

	Proposed algorithm	Multi-operator	Cropping	Streaming video
Mean rank	2.44	2.53	2.18	2.83
Standard deviation	1.24	1.16	1.07	1.11

An example of the no-reference comparison is shown in Figure 5.

As is shown in [8], manually cropping always achieves the best result in no-reference comparison case since the cropping window is picked manually and no deformation is inserted. But manually cropping is unrealizable in most cases. The above results show that our proposed algorithm makes a better balance between content loss and deformation, and achieves better retargeting result than the other 2 operators.

Technique in [9] also uses the idea of strip process in media retargeting. A comparison example between technique in [9] and the proposed approach can be seen in Figure 6. As the proposed algorithm considers both structure information and deformation and combines several opeators for retargeting, it can obviously improve the result.

Moreover, the computational cost of the proposed algorithm is low. Most of the computational complexity of the the proposed scheme comes from the energy function acquisition and strip retargeting steps. As discussed above, these steps can

(a) Multi-operator

(b) Cropping

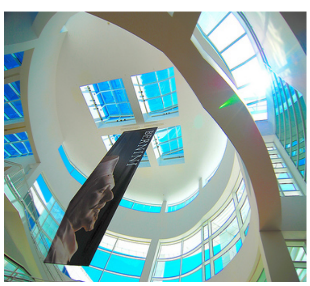
(c) Streaming video

(d) Proposed algorithm

Fig. 5. An example of no-reference retargeting. (a)-(d) are the result applying four retargeting operators, respectively. The result image is reduced to 75% in width of the original image.

(a) Original image

(b) Result via proposed algorithm



(c) Result via method in [9]

Fig. 6. A comparison example of the proposed algorithm and method in [9]. The width of the result image is resized to 75% of the source image. It can be seen that the proposed algorithm can make better visual friendly result.

be achieved in real time, that means the proposed retargeting scheme is real time achievable.

V. Conclusion

We proposed a strip based algorithm for media retargeting via combing multi-operators. This algorithm makes a good balance between content loss and deformation, and achieves better performance than most state-of-the-art media retargeting operators. Moreover, the computational cost of the proposed algorithm is low and can be done realtimely.

References

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing" in *ACM Trans. Graphics*, 26(3), July 2007.

[2] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting" in *ACM Trans. Graphics*, 27(3), Aug. 2008.

[3] A. Mansfield, P. Gehler, L. V. Gool and C. Rother, "Visibility Maps for Improving Seam Carving" in *Media Retargeting Workshop, European Conference on Computer Vision (ECCV)*, 2010.

[4] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video" in *ACM Trans. Graphics*, 28(5):1-10, 2009.

[5] L. Wolf, M. Guttmann, D. Cohen-Or, "Nonhomogeneous content-driven video-retargeting" in *ICCV*, 2007.

[6] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing" in *ACM Trans. Graphics*, 27(5), 2008.

[7] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel, "Motion-aware temporal coherence for video resizing" in *ACM Trans. Graphics*, 28(5), 2009.

[8] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting" in *ACM SIGGRAPH Asia 2010 papers* , 160:10, ACM, New York, NY, USA, 2010.

[9] J.-S. Kim, J.-H. Kim, and C.-S. Kim, "Adaptive image and video retargeting technique based on Fourier analysis" in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, pp. 1730-1737, Sept. 2009.

[10] L. Itti, C. Koch, and E. Niebur, "A model of saliencybased visual attention for rapid scene analysis" in *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[11] P. Viola and M. J. Jones, "Robust real-time face detection" in *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137-154, 2004.

[12] H. Huang, T. Fu, P. Rosin, and C. Qi, "Real-time contentaware image resizing" in *Science in China Series F: Information Sciences*, vol.52, no. 2, pp. 172-182, 2009.

[13] C. Guo, Q. Ma and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform" in *CVPR*, 2008.

[14] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting" in *ACM Trans. Graphics*, 28(3):1-11, 2009.

[15] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach" in *CVPR*, 2007.