SALIENCY DETECTION FOR CONTENT-AWARE IMAGE RESIZING

Radhakrishna Achanta and Sabine Süsstrunk

School of Computer and Communication Sciences (IC)
Ecole Polytechnique Fédérale de Lausanne (EPFL)

ABSTRACT

Content aware image re-targeting methods aim to arbitrarily change image aspect ratios while preserving visually prominent features. To determine visual importance of pixels, existing re-targeting schemes mostly rely on grayscale intensity gradient maps. These maps show higher energy only at edges of objects, are sensitive to noise, and may result in deforming salient objects. In this paper, we present a computationally efficient, noise robust re-targeting scheme based on seam carving by using saliency maps that assign higher importance to visually prominent whole regions (and not just edges). This is achieved by computing global saliency of pixels using intensity as well as color features. Our saliency maps easily avoid artifacts that conventional seam carving generates and are more robust in the presence of noise. Also, unlike gradient maps, which may have to be recomputed several times during a seam carving based re-targeting operation, our saliency maps are computed only once independent of the number of seams added or removed.

Index Terms— Content aware image re-targeting, seam carving, saliency.

顯著的

1. INTRODUCTION

The diversity of today's display device sizes and aspect ratios demands smarter ways of re-targeting images than simple resizing to better deliver visually important or salient content for the given display dimensions. While cropping [1] is one option, image content adaptive warping [2] and seam carving [3] are methods that accentuate visually important content with minimal loss of original intent. These two re-targeting approaches have also been extended to videos [4, 5].

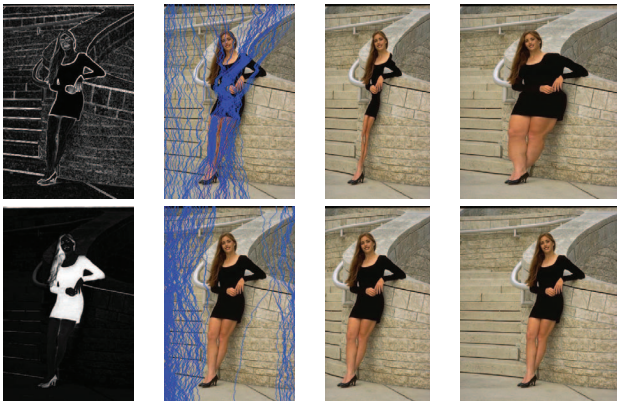
Gal et al. [6] were the first to propose a solution to the general problem of re-targeting an image while preserving regions of interest. In their method, the user has to manually specify the regions of interest based on which the image is adaptively warped.

Automatic *content awareness*, i.e the choice of visually important regions in re-targeting schemes, was introduced by [2, 3, 4, 5]. All such automatic re-targeting methods rely on finding visual importance values for each pixel. Avidan and Shamir [3] proposed the popular content aware re-targeting scheme of *seam carving*. They iteratively remove a seam, i.e. a connected set of vertical (horizontal) pixels, to reduce the width (height) of an image.

Rubinstein et al. [4] extend the original seam carving idea of [3] to videos by removing pixel manifolds in 3D volumes of video frames. Video re-targeting is also done by Wolf et al. [5] who locally warp the frames of the video. This is done by optimizing the best mapping between a source image and the re-targeted image.

Wang et al. [2] present another way of re-targeting images to arbitrary aspect ratios while preserving visually prominent features given by a saliency map using mesh based adaptive warping.

Assigning visual importance values (Section 2) is fundamental to all these automatic re-targeting methods. In this paper, we present a method of generating a saliency map that assigns visual importance to each pixel in terms of its global color and intensity contrast. We demonstrate that our saliency maps [7] provide better seam carving results than the usual gradient maps (see Fig. 1).



(A) (B) (C) (D)

Fig. 1. *Our seam carving compared to state-of-the-art [4]: top row results show the use of intensity gradient based energy map [4]; bottom row results show the use of our global contrast based saliency map. Column (A) is the visual importance map (the darker the pixel the lower the importance). Column (B) shows the seams chosen for re-targeting, superimposed on the original image. Column (C) shows image with 80% width. Column (D) shows image with 120% width.*

2. VISUAL IMPORTANCE MAPS

"saliency map"
"energy map"

The key to content awareness, needed by all automatic re-targeting schemes, is a map of values that quantifies the relative visual importance of each pixel. The main methods of assigning importance values to pixels are measures of L_1-norm [3, 4] or L_2 norm [5] of the grayscale intensity gradient, face or other object detectors, saliency maps, or a combination of these [2, 5].

Avidan and Shamir [3] use the L_1-norm of the grayscale intensity gradient to compute their energy map. The energy map lets them

This work is supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322, and the European Commission under contract FP6-027026 (K-Space, the European Network of Excellence in Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content).

successively remove seams of minimal energy as determined using a dynamic programming algorithm. They compare several ways of computing the energy maps, including Itti's saliency maps [8], and conclude that sums of magnitudes of gradients along the x and y axes (Eq. 1) and the same normalized by the maximum of the histogram of oriented gradients [9] give good results in general. To extend the spatial energy computation of Avidan and Shamir [3] to a spatiotemporal one for their video re-targeting case, Rubinstein et al. [4] introduce an inter-frame L_1-norm gradient term.

Wolf et al. [5] use a saliency map that combines the result of a face detector and a motion detector with the L_2-norm of the intensity gradient. The importance map of Wang et al. [2] is generated by multiplying the L_2-norm of the intensity gradient of the image with Itti's saliency maps [8]. Itti's maps do not highlight salient regions uniformly and are highly downsized as compared to the input image [10]. Thus, the resulting energy map has lower values for gradients that are not in the vicinity of a saliency blob of Itti's map.

Pixel energy computed from simple L_1-norm [3, 4] or L_2-norm [5] of the grayscale intensity gradient suffers from certain drawbacks. First, the values peak at edges rather than whole salient regions. Thus, energy is assigned to visually important image content only at edges and not whole regions (see Fig. 1, top-left image). Second, color information is ignored. The third disadvantage w.r.t iterative re-targeting schemes like seam carving [3] is the need to recompute the energy after seams are removed, since the local gradients may change after a seam is removed. Finally, gradient based maps can be noise sensitive.

Our saliency maps [7] uniformly assign saliency values to entire salient regions, rather than just edges or texture regions. This is achieved by relying on the global contrast of a pixel rather than local contrast, measured in terms of both color and intensity features rather than just intensity as done previously [3, 4]. The saliency map is computed only once irrespective of the number of seam carving operations performed and is robust in the presence of noise. We show the effectiveness of our method in avoiding the usual artifacts of seam carving in normal and noisy images.

3. SEAM CARVING REVIEW

Avidan and Shamir [3] define a vertical (horizontal) seam to be an 8-connected path of low energy pixels in the image from top to bottom (left to right) containing one, and only one, pixel in each row (column) of the image. Thus, removing a vertical (horizontal) seam reduces the width (height) by one pixel. Finding the globally minimum energy seam, which removes the least salient content, is posed as a dynamic programming optimization problem. The energy maps are computed using the L_1-norm of the intensity gradient as:

gradient: 梯度

$$E_g(x,y) = \left|\frac{\partial}{\partial x}I(x,y)\right| + \left|\frac{\partial}{\partial y}I(x,y)\right| \tag{1}$$

where $E_g(x,y)$ is the resulting importance value of a pixel at column x and row y, and I is the grayscale intensity image. For a vertical seam removal, the dynamic programming memoization table entry $M(x,y)$ is given as:

L1 norm: 曼哈頓距離
L2 norm: 歐氏距離

$$M(x,y) = E_g(x,y) + min\begin{cases} M(x-1,y-1) \\ M(x,y-1) \\ M(x+1,y-1) \end{cases} \tag{2}$$

https://medium.com/@montjoile/l0-norm-l1-norm-l2-norm-l-infinity-norm-7a7d18a4f40c

The globally minimum energy seam is found by backtracking from the minimum value of the last row in M to the first row.

Rubinstein et al. [4] note that despite seam carving being an energy removal operation, a removed seam may actually introduce more energy than it takes away because of previously non-adjacent pixels becoming neighbors. They therefore introduce *forward energy* criteria (equally applicable for both image and video cases; refer to [4] for details), such that the optimal seam found is one whose removal re-introduces minimum amount of energy. This changes Eq. 2 to:

$$M(x,y) = E_g(x,y) + min\begin{cases} C_L(x,y) + M(x-1,y-1) \\ C_U(x,y) + M(x,y-1) \\ C_R(x,y) + M(x+1,y-1) \end{cases} \tag{3}$$

where C_L, C_U, and C_R are image gradients resulting from non-adjacent pixels becoming neighbors when a seam pixel separating them is removed, and are computed as:

$$\begin{aligned} C_U(x,y) &= |I(x+1,y) - I(x-1,y)| \\ C_L(x,y) &= |I(x,y-1) - I(x-1,y)| + C_U(x,y) \\ C_R(x,y) &= |I(x,y-1) - I(x+1,y)| + C_U(x,y) \end{aligned} \tag{4}$$

4. OUR SALIENCY MAP

As mentioned above, the importance maps used by [2, 3, 4, 5] determine local grayscale contrast using gradients that result in higher importance values for textured areas and edges, but lower values for smooth salient regions. Wang et al. [2] attempt to address this problem by multiplying the L_2 norm of the gradient with Itti's saliency maps [8], while Wolf et al. [5] combine the result of a face detector and a motion detector with the L_2-norm of the intensity gradient. However, these do not significantly alleviate the drawbacks of intensity gradient maps.

Our saliency map [7] is obtained by evaluating the Euclidean distance of the average Lab vector value of an input image with each pixel of a Gaussian blurred version (using a 3×3 or 5×5 binomial kernel) of the same input image:

$$E_{Lab}(x,y) = \|\mathbf{I}_\mu - \mathbf{I}_{n\times n}(x,y)\| \tag{5}$$

where $E_{lab}(x,y)$ is the pixel importance value at position (x,y), $\mathbf{I}_\mu$ is the average of all Lab pixel vectors of the image, $\mathbf{I}_{n\times n}(x,y)$ is the corresponding image pixel vector value in the Gaussian blurred version of the original image, and $\|\|$ is the L_2 norm (i.e. Euclidean distance in Lab color space). We use the Lab color space since Euclidean distances in this color space are approximately perceptually uniform. Our saliency maps have uniformly highlighted salient regions with well-defined boundaries, an improvement over several state-of-the-art methods [7].

We also introduce the use of color information in the forward energy terms by replacing the scalar gray scale differences in Eq. 4 with the corresponding vector distances in Lab space to obtain:

$$\begin{aligned} C_U(x,y) &= \|\mathbf{I}(x+1,y) - \mathbf{I}(x-1,y)\| \\ C_L(x,y) &= \|\mathbf{I}(x,y-1) - \mathbf{I}(x-1,y)\| + C_U(x,y) \\ C_R(x,y) &= \|\mathbf{I}(x,y-1) - \mathbf{I}(x+1,y)\| + C_U(x,y) \end{aligned} \tag{6}$$

This computes forward energy better as both color and intensity information is taken into account. Our saliency maps (see Figs. 1, and 2) generated using Eq. 5 coupled with the modified forward energy terms (Eq. 6) overcome the limitations of importance maps used previously by re-targeting schemes [2, 3, 4, 5] of Section 2. Since only an averaging operation and a Gaussian blurring (with separable

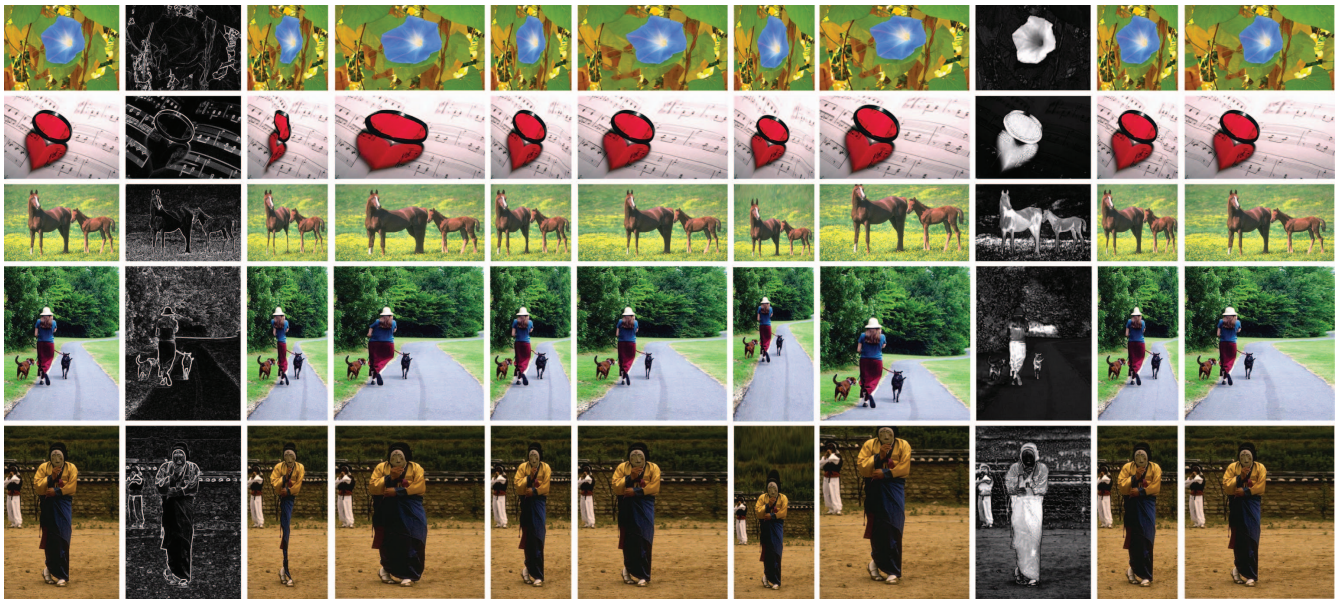


Fig. 2. *Image re-targeting comparison for 70% and 130% of original width: Column(A) is the original image, (B) is the gradient map from Eq. 1, (C) and (D) are the outputs of [4], (E) and (F) are the outputs of [5], (G) and (H) are the outputs of [2], (I) is our saliency map, and (J) and (K) are our seam carving results.*

binomial kernel) over three channels is required, the complexity is $O(N)$, N being the number of pixels in the input image, and the saliency maps can be computed in real time.

5. IMPROVED SEAM CARVING

The seam carving results[1] obtained by using our saliency map (Eq. 5) and modified forward energy terms (Eq. 6) are compared against those of [4] obtained using gradient maps (Fig. 2B), as well as those of [2] and [5] in Fig. 2.

We show results for changing aspect ratios by both removing and adding seems as needed. As proposed by Avidan and Shamir [3], to enlarge an image by k seams, we first find k seams for removal and duplicate them. To affect a change in both dimensions of an input image, we choose between a vertical or a horizontal seam at each step depending on which has lower energy.

Our saliency maps highlight visually important regions of the image uniformly and not just edges. Thus, the seams chosen do not pass through high energy regions, such as salient objects (see Fig. 1B). This permits us to obtain seam carving results without artifacts in salient regions. It must be added, in cases where the salient object is not highlighted correctly by our maps, gradient based energy maps from Eq. 1 may provide better re-targeting results.

In our saliency maps, the importance value associated with a pixel is computed with respect to the entire image (and not the immediate four or eight neighbors). High saliency values are not assigned just at the edge of a region, but the entire region. Once we know which pixels are less salient with respect to the original image, we can remove them without having to recompute their importance after each removal. This is unlike local gradients, whose values depend on local pixel neighborhood, which may change when a seam of pixels is removed.

[1]The images used in our work are from the Berkeley database (www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/) and the MSRA salient object database (http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient_object.htm). More results at: http://ivrg.epfl.ch/supplementary_material/RK_ICIP09/index.html

6. NOISE ROBUSTNESS

Our saliency maps are more robust to noise than local intensity gradient based maps. There are two reasons for this. First, our global approach is independent of local noise patterns that strongly affect gradient based energy maps. Second, Eq. 5 allows using Gaussian blurring that can be increased according to the requirements of the application. Although a 3×3 binomial kernel suffices, if very low bit-rate coding is used or if Exif (Exchangeable Image File Format [11]) data indicates the use of high ISO values (indicating probability for higher noise), one can increase the binomial kernel size.

We experiment with Gaussian noise up to variance 0.1 (Fig. 3), and salt and pepper noise with noise density up to 0.1. We retain the binomial kernel at 3×3 for all experiments (with and without noise). Our saliency maps provide good seam carving results even in the presence of noise, as illustrated in Fig. 3.

7. CONCLUSIONS

We present an improved method of re-targeting images that uses a novel saliency detection scheme, which is easy to implement, computationally inexpensive, and has the same resolution as the original image. We demonstrate the efficacy of the saliency maps in re-targeting images using seam carving where most artifacts arising from conventional intensity gradient based energy maps are easily avoided. There are three advantages of using our saliency maps for any re-targeting scheme: they exploit both color and intensity information of the image, they are computed only once irrespective of the

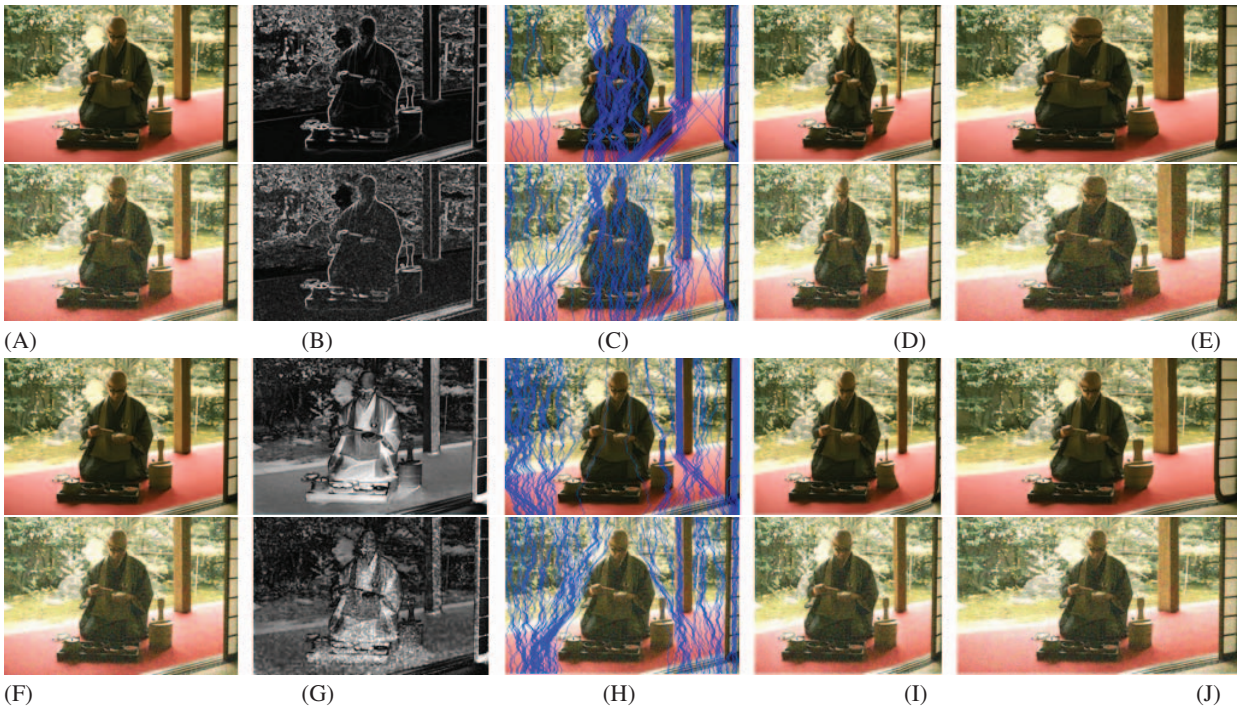


Fig. 3. *Column (A) shows original (above) and noisy (below) versions of the image. Column (B) shows the corresponding gradient energy maps. Column (C) shows the seams chosen based on the gradient maps. Columns (D) and (E) show width re-targeted to 80% and 120% of the original. Columns (F) to (J) show the result of using our maps instead. Note how much the seam selection is affected by noise in gradient maps as compared to our maps. Despite the noise, the seams continue to be chosen from the same regions of the image (Column (H)).*

number of seam carving operations performed, and they allow good seam carving even in the presence of noise.

8. ACKNOWLEDGEMENTS

We thank Yu-Shuen Wang and Tong-Yee Lee (authors of [2]), and Lior Wolf and Moshe Guttmann (authors of [5]) for providing their image re-targeting results for the comparison presented in Fig. 2.

9. REFERENCES

[1] Liqun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and Heqin Zhou, "A visual attention model for adapting images on small displays," *ACM Transactions on Multimedia Systems*, vol. 9, pp. 353–364, November 2003.

[2] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkin, and Tong-Yee Lee, "Optimized scale-and-stretch for image resizing," *ACM Transactions on Graphics (Also Proceedings of ACM SIGGRAPH ASIA)*, vol. 27, no. 5, 2008.

[3] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 26, no. 3, pp. 10, July 2007.

[4] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Improved seam carving for video retargeting," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 27, 2008.

[5] Lior Wolf, Moshe Guttmann, and Daniel Cohen-Or, "Non-homogeneous content-driven video-retargeting," *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–6, Oct. 2007.

[6] Ran Gal, Olga Sorkine, and Daniel Cohen-Or, "Feature-aware texturing," *Eurographics Symposium on Rendering*, pp. 297–303, 2006.

[7] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Süsstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1597–1604, June 2009.

[8] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[9] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893 vol. 1, June 2005.

[10] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk, "Salient region detection and segmentation," *International Conference on Computer Vision Systems (ICVS)*, vol. 5008, pp. 66–75, 2008.

[11] "Digital still camera image file format standard (exchangeable image file format for digital still cameras: Exif) Version 2.1, Specification by JEITA," June 1998.