PERCEPTUALLY RELEVANT ENERGY FUNCTION FOR SEAM CARVING

Hui Li Tan[1,2], *Yih Han Tan*[1], *Zhengguo Li*[1], *Susanto Rahardja*[1,2], *Chuohao Yeo*[1]

{hltan, yhtan, ezgli, rsusanto, chyeo}@i2r.a-star.edu.sg
[1] Signal Processing Department, Institute for Infocomm Research, A*STAR
[2] Department of Electrical and Computer Engineering, National University of Singapore

ABSTRACT

Seam carving, an image re-targeting method, works by progressively finding and removing connected paths of low energy pixels in an image until a desired image aspect ratio is reached. In this paper, we first cast the problem of minimizing an energy function as that of minimizing a distortion cost. We then leverage on our understanding of image quality metrics/distortion metrics in proposing a perceptually relevant energy function. Experimental results show that our proposed energy function can generate more desirable resized images in which the original structures of the images are better preserved.

Index Terms— Seam carving, context-aware image resizing, image re-targeting, perceptual image quality metrics, distortion metrics.

1. INTRODUCTION

Digital images are often viewed on different display devices with different resolutions. Hence, images have to be resized according to some image aspect ratios to accommodate these different applications. This is known as image re-targetting. Naive image re-targetting methods such as cropping and scaling often reduce important content of an image and/or introduce perceivable distortions. Recognizing the need for more advanced image re-targetting methods, researchers in recent years developed methods such as seam carving [1], shift-map editing (graph-cut based method) [2] and warping [3], which attempt to resize images while preserving their content information and salient shapes.

Seam carving (also known as content-aware image resizing) works by first finding connected paths of low energy pixels in an image and then progressively removing them until a desired image aspect ratio is reached. These connected paths of low energy pixels are known as seams.

For an image of size $W \times H$, a vertical seam is defined as:

$$\mathbf{s^v} = \{s_y\}_{y=1}^{H} = \{(f_x(y), y)\}_{y=1}^{H}, \tag{1}$$

where f_x is a mapping $f_x : [1, \ldots, H] \to [1, \ldots, W]$ such that $|f_x(y) - f_x(y-1)| \le 1$ for all $y > 1$. Similarly, a horizontal seam is defined as:

$$\mathbf{s^h} = \{s_x\}_{x=1}^{W} = \{(x, f_y(x))\}_{x=1}^{W}, \tag{2}$$

where f_y is a mapping $f_y : [1, \ldots, W] \to [1, \ldots, H]$ such that $|f_y(x) - f_y(x-1)| \le 1$ for all $x > 1$. Without loss of generality, we will take $\mathbf{s} = \mathbf{s^v}$, and illustrate the following procedures for the removal of vertical seams.

In each step, the optimal seam to be removed, $\mathbf{s}^*$, is one that minimizes the seam cost, i.e.,

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} E(\mathbf{s}) = \arg\min_{\mathbf{s}} \sum_{y=1}^{H} e(s_y), \tag{3}$$

where $e(\bullet)$ is an energy function. The minimization of the energy path in an image can be achieved by using Dijkstra's algorithm, which is a dynamic programming method. That is, for $1 < x < W$, the cumulative energy $M(x, y)$ is found by:

$$M(x,y) = \begin{cases} e(x,y) & y = 1; \\ e(x,y) + \min\{M(x-1,y-1), & \\ M(x,y-1), M(x+1,y-1)\} & 2 \le y \le H. \end{cases} \tag{4}$$

The minimum value of the last row in $M(x, y)$, i.e., $\min(M(x, H))$, indicates the end of the optimal connected vertical seam and backtracking on this minimum entry gives the path of the optimal seam $\mathbf{s}^*$.

Clearly, depending on the definition of the energy function, the optimal seam found and removed would be different. In this paper, we first cast the problem of finding an appropriate energy function as that of finding an appropriate image quality metric/distortion metric. We then leverage on our understanding of perceptual image quality metrics to propose a perceptually relevant energy function.

The paper is organized as follows. In Section 2, a review of previously proposed energy functions is presented. Subsequently, a perceptually relevant energy function is proposed in Section 3 and experimental results to demonstrate the performance of the proposed energy function are presented in Section 4. Finally, in Section 5, the conclusion and our future works are listed.

2. RELATED WORKS

Several energy functions have been explored in previous works. In the seminal work [1], three energy functions are studied, i.e.,

$$e_1(x,y) = \left|\frac{d}{dx}I(x,y)\right| + \left|\frac{d}{dy}I(x,y)\right|, \quad (5)$$

$$e_2(x,y) = \frac{d^2}{dx}I(x,y) + \frac{d^2}{dy}I(x,y), \text{ and} \quad (6)$$

$$e_{HOG}(x,y) = \frac{e_1(x,y)}{\max(HOG_{nbhd(x,y)})}, \quad (7)$$

where $I(x,y)$ denotes the intensity of pixel (x,y). e_1 is the absolute sum of the partial derivatives in the horizontal and vertical directions, e_2 is the sum of the second order partial derivatives in the horizontal and vertical directions, and e_{HOG} is e_1 normalized by the maximum of the histogram of gradients measure [4] computed in a neighbourhood about the pixel (x,y). The aim of e_{HOG} is to attract a seam to an edge but not cross it. Their experiments showed that no single energy function performs well across the range of images but e_1 and e_{HOG} work reasonably well in general.

Recognizing that the decomposition of images into different frequency sub-bands would provide a more effective representation of image features, Han et al. proposed a modified energy function based on wavelet decomposition [5], i.e.,

$$\begin{aligned} & e_{wavelet}(x,y) \\ = \; & e_1(x,y) + \sum_{l=1}^{L} \Phi_l(x,y) \quad (8) \\ = \; & e_1(x,y) + \\ & \sum_{l=1}^{L} \alpha\omega_{LH}^l(x,y) + (1-\alpha)\omega_{HL}^l(x,y) + \beta\omega_{HH}^l(x,y), \quad (9) \end{aligned}$$

where $\omega_{LH}^l(x,y)$, $\omega_{HL}^l(x,y)$, $\omega_{HH}^l(x,y)$ are the wavelet coefficients for each sub-band in the l^{th} level. $L = 3$, α, β are empirically determined parameters. For vertical seam carving, $\alpha = 0.2$, $\beta = 0.25$ [5], so that more vertical than horizontal edges are preserved.

Similarly, Hwang et al. proposed another modified energy function by considering the human attention model [6], i.e.,

$$e_{fs}(x,y) = e_1(x,y) + w_f F(x,y) + w_s S(x,y), \quad (10)$$

where $F(x,y)$ and $S(x,y)$ are values from a face map [7] and a saliency map [8] respectively. $w_f = w_s = 0.30$ are empirically determined. Their proposed energy function is shown to be effective in preserving facial features.

3. PROPOSED METHOD

Let $I(x,y)$ be the intensity of the pixel at (x,y). Then $I(x-1,y)$ and $I(x,y-1)$ are the intensities of the pixels to the left of and above the pixel at (x,y) respectively. Since seam carving can be viewed as removing optimal seam s* and replacing the intensities of that seam with the intensities of the immediate neighbouring pixels,

$$\left|\frac{d}{dx}I(x,y)\right| \approx |I(x,y) - I(x-1,y)|, \text{ and} \quad (11)$$

$$\left|\frac{d}{dy}I(x,y)\right| \approx |I(x,y) - I(x,y-1)|, \quad (12)$$

can be interpreted as the distortion cost of replacing the current pixel with that of its immediate left and above neighbouring pixels. In other words, the problem of minimizing an energy function can be cast as the problem of minimizing a distortion cost between two neighbouring pixels. We can now utilize knowledge in the studies of perceptually relevant distortion metrics in designing a perceptually relevant energy function.

In [10], [11], a perceptually relevant image quality metric was proposed and we define it here as:

$$\text{M-MSE-SSIM}(I, \tilde{I}) = \frac{1}{M}\sum_{j=1}^{M} \text{MSE-SSIM}(I_j, \tilde{I}_j), \quad (13)$$

where I and $\tilde{I}$ denote the pixel intensities of a source image and its distorted image respectively; I_j and $\tilde{I}_j$ are the pixel intensities at the corresponding j^{th} local window of I and $\tilde{I}$ respectively.

$$\text{MSE-SSIM}(I_j, \tilde{I}_j) = \frac{2\sigma_{I_j}^2 + c_2}{2\sigma_{I_j}^2 + \text{MSE}(I_j, \tilde{I}_j) + c_2}, \quad (14)$$

where $\text{MSE}(I_j, \tilde{I}_j)$ is the mean-squared-error (MSE) between I_j and $\tilde{I}_j$, and $\sigma_{I_j}^2$ is the variance of I_j. c_2 is a constant used for numerical stability. This theoretically derived image quality metric has been empirically shown to be well-correlated to subjective visual quality [11].

The proposed objective image quality metric measures the fidelity of a distorted image to its source image. To measure the distortion between two images, we can consider:

$$\begin{aligned} \text{d-MSE-SSIM}(I_j, \tilde{I}_j) &= \frac{1}{\text{MSE-SSIM}(I_j, \tilde{I}_j)} - 1 \\ &= \frac{\text{MSE}(I_j, \tilde{I}_j)}{2\sigma_{I_j}^2 + c_2}. \quad (15) \end{aligned}$$

(15) is simply a weighted MSE, being normalized by its local variance.

Since $\left|\frac{d}{dx}I(x,y)\right|^2$ and $\left|\frac{d}{dy}I(x,y)\right|^2$ measures the local distortion at (x,y) if the seam passes through it, we propose to similarly normalise them by the local variance, i.e.,

$$\sqrt{\frac{\left|\frac{d}{dx}I(x,y)\right|^2}{2\sigma^2_{I_{(x,y)}}+c_2}}+\sqrt{\frac{\left|\frac{d}{dy}I(x,y)\right|^2}{2\sigma^2_{I_{(x,y)}}+c_2}} \quad (16)$$

$$= \frac{\left|\frac{d}{dx}I(x,y)\right|+\left|\frac{d}{dy}I(x,y)\right|}{\sqrt{2\sigma^2_{I_{(x,y)}}+c_2}}, \quad (17)$$

where $\sigma^2_{I_{(x,y)}}$ is the local variance computed in a neighbourhood about the pixel at (x, y).

We thus propose a modified energy function as:

$$e_{proposed}(x,y) = \frac{\left|\frac{d}{dx}I(x,y)\right|+\left|\frac{d}{dy}I(x,y)\right|}{\sqrt{2\sigma^2_{I_{(x,y)}}+c_2}}. \quad (18)$$

It is commonly known that the human visual system (HVS) allows more distortions in a textured region than a smooth region. Hence, seams removed from a textured region will be less noticeable than seams removed from a smooth region. Our proposed energy function is consistent with this property. A region with a smaller local variance will be subjected to less removal than a region with a larger local variance.

4. EXPERIMENTAL RESULTS

We evaluated the performance of our proposed energy function on test images from an image re-targeting database [13]. In our evaluations, the test images are subjected to 50% reduction in image width. Three energy function were evaluated, namely, e_1, e_{HOG}, and $e_{proposed}$. 9×9 cell size was used for e_{HOG} and 9×9 block size, without overlap, was used for $e_{proposed}$.

The results of some test images are shown in Figures 1 to 3. It can be seen that there is better preservation of geometric shapes and structures in images seam carved using the proposed energy function. This can be credited to the distinction between the significance of edges in smooth and textured regions. When e_1 is used, edges are considered to be more important than non-edges and hence will be preserved to a greater extent. Hence, when a significant number of seams are removed, there might be an imbalance in the proportion of edges and non-edges, leading to geometric distortions. On the other hand, $e_{proposed}$ takes into consideration local variance to perceptually distinguish the significance of the edges when removed.

We also evaluated the performance of energy functions using the Earth mover's distances (EMD) [14], which has been studied as a possible metric for measuring the distortion in object shapes of re-targeted images [15]. The EMD of two identical images is zero. Also shown in Figures 1 to 3, the EMDs of the original images and images seamed carved with the proposed energy function are the lowest.

The run-times for the computation of the three energy functions for a test image are also listed in Table 1. Evaluations were performed on a PC with 2.66GHz Intel Core 2 CPU and 24GB of RAM. MATLAB implementations were used in all evaluations. Since e_{HOG} and $e_{proposed}$ require computations of local HOG features and variance in a neighbourhood about the pixels, their run-times are higher than that of e_1.

Table 1. Run-times for the computation of the energy functions for 'Buddha.png'.

Energy Function	Computational Time (s)
e_1	0.132
e_{HOG}	9.142
$e_{proposed}$	0.437

5. CONCLUSION

Seam carving works by progressively finding and removing connected paths of low energy pixels in an image until a desired image aspect ratio is reached. In this paper, we illustrate how to interpret the problem of minimizing an appropriate energy function as that minimizing an appropriate distortion metric. Thereby, we leverage on our understanding of perceptually relevant image quality metrics/distortion metrics in proposing a perceptually relevant energy function. Experimental results show that our proposed energy function can generate more desirable resized images where the original geometric shapes and structures of the images are better preserved.

This paper is an effort to bring the knowledge associated with perceptual image quality metrics into seam carving, to illustrate the use of the image quality metrics/distortion metrics in image processing applications. Comparisons with the wavelet based method and the face map and saliency map method would constitute our future works. Nonetheless, our proposed energy function is not mutually exclusive with those methods and it is conjectured that fusion with those methods would lead to more desirable image re-targeting results.

6. REFERENCES

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.

[2] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Sep. 2009, pp. 151–158.

(a) (b) (c) (d)

Fig. 1. 'Brasserie_L_Aficion.png' (a) original image (b) seam carved using e_1 (EMD = 0.028) (c) seam carved using e_{HOG} (EMD = 0.026) (d) seam carved using $e_{proposed}$ (EMD = 0.008).

(a) (b) (c) (d)

Fig. 2. 'buddha.png' (a) original image (b) seam carved using e_1 (EMD = 0.066) (c) seam carved using e_{HOG} (EMD = 0.067) (d) seam carved using $e_{proposed}$ (EMD = 0.006).

(a) (b) (c) (d)

Fig. 3. 'Marblehead_Mass.png' (a) original image (b) seam carved using e_1 (EMD = 0.020) (c) seam carved using e_{HOG} (EMD = 0.016) (d) seam carved using $e_{proposed}$ (EMD = 0.009).

[3] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV)*, 2007.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, Jun. 2005, pp. 886–893.

[5] J.-W. Han, K.-S. Choi, T.-S. Wang, S.-H. Cheon, and S.-J. Ko, "Improved seam carving using a modified energy function based on wavelet decomposition," in *IEEE 13th International Symposium on Consumer Electronics (ISCE)*, May 2009, pp. 38–41.

[6] D.-S. Hwang and S.-Y. Chien, "Content-aware image resizing using perceptual seam carving with human attention model," in *IEEE International Conference on Multimedia and Expo*, 23-26 Apr. 2008, pp. 1029–1032.

[7] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Trans. Appl. Percept.*, vol. 7, no. 1, pp. 6:1–6:39, Jan. 2010.

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

[9] C. Yeo, H. L. Tan, and Y. H. Tan, "On rate distortion optimization using SSIM," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 833–836.

[10] C. Yeo, H.-L. Tan, and Y. H. Tan, "On rate distortion optimization using SSIM," *To be published in IEEE Transactions on Circuits and Systems for Video Technology*.

[11] H. L. Tan, Z. Li, Y. H. Tan, S. Rahardja, and C. Yeo, "A perceptually relevant image quality metric," *Submitted to IEEE Transactions on Image Processing*.

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[13] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, vol. 29, no. 5, pp. 160:1–160:10, 2010.

[14] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Computer Vision, 2009 IEEE 12th International Conference on*, 29 2009-oct. 2 2009, pp. 460 –467.

[15] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 626–639, Oct. 2012.