Neurocomputing 151 (2015) 645–653

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Improved seam carving combining with 3D saliency for image retargeting

Yanxiang Chen *, Yifei Pan, Minglong Song, Meng Wang

School of Computer & Information, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 14 November 2013
Received in revised form
7 March 2014
Accepted 17 May 2014
Communicated by Chennai Guest Editor
Available online 23 October 2014

Keywords:
Seam carving
3D saliency
Image retargeting
Gradient
Depth
Energy map

ABSTRACT

Seam carving is a content-aware image retargeting algorithm that removes the pixels with less energy values during resizing process to preserve important parts. Though many researchers have improved this algorithm by different ways, it is still a difficult problem to determine the energy function for removing task. Most existing approaches only use 2D features, and no single energy function performs well across all kinds of images so far. In this paper, we introduce a novel method that combine conventional $L-1$ norm of gradient with depth-aware saliency (3D saliency) to obtain energy map. Due to the different characteristics of these two operators in image analysis, our energy function contains both local and global information, which is proven to be effective in seam carving. The experimental results demonstrate the advantage of our method compared to conventional seam carving techniques.

1. Introduction

As mobile devices such as cell phones, tablet PCs are becoming increasingly popular, people would like to share their moments with their friends or relatives. They will use mobile terminal to upload photos to social networks, and their friends also get these data through the mobile portable displays. How to display the image on different terminal screens without distortion has become increasingly urgent need, but there are still many challenges to be accomplished.

Lots of papers have devoted to solving this problem, and some classical methods are available, such as scaling, cropping, etc. Scaling is the most direct and least desirable method. It achieves different resolutions display through stretching and interpolation, which will lead to a serious distortion of image content and structure.

Cropping is another traditional method, in which the most important step is to establish saliency map [1,2] and find the ROI (Region of Interest) of images, and then directly cropping the corresponding portion of the ROI according to different display screens [3,4]. But this will result in the loss of a large number of image backgrounds, and the results are not satisfactory when there are many ROIs in the source image. An approach [5] has been proposed to solve the problem, especially when the set of ROIs cannot contain all the important regions. They remove the important regions from the image and the resulting holes are filled using inpainting. Then, the background is resized to fit the display specification, and the important regions are pasted back onto the updated background. This approach also results in the distortion of the background. In short, cropping images to fit the different display mediums inevitably discards information, including backgrounds, structure, etc.

Recently, content-aware methods such as non-uniform warping [6,7] and seam carving [8] are proposed to supplement these traditional methods. Content-aware image retargeting achieves to change image into arbitrary aspect ratios while preserving visually prominent features, such as image structure. It relies on an importance map to retain the important parts of the picture at the expense of these less-important parts. Generally, the importance map (or energy function) can be generated by image gradients, saliency or entropy.

Warping is a special method of content ware methods. It puts a grid mesh onto the original image and then computes a new geometry for this mesh, so that boundaries adapt to the new size of the desired image and the quad faces covering important image regions remain unchanged at the expense of larger distortion to the other quads. Unfortunately, this method will fail if the size of prominent objects is larger than the desired image size.

Seam carving is a simple image operator which supports content-aware image resizing for both reduction and expansion. It delimits the importance of pixels using an energy function. A seam is an optimal 8-connected path of pixels on a single image from top to bottom, or left to right. By repeatedly carving out or inserting seams in one direction, the aspect ratio of an image can

* Corresponding author. Tel.: +86 13966668160.
E-mail address: chenyx@hfut.edu.cn (Y. Chen).

http://dx.doi.org/10.1016/j.neucom.2014.05.089
0925-2312/

be changed. By applying these operators in both directions, the image can be retargeted to different sizes. Compared to the conventional methods, seam carving has great advantages. For image reduction, seam selection ensures to preserve the image structure by removing more low-energy pixels and fewer high-energy ones. For image enlarging, the order of seam insertion ensures a balance between the inserted pixels and the original image content. However, the grayscale intensity gradient maps generating the energy function have higher value only at edges of objects and are sensitive to noise, which may result in deforming the salient objects in the image.

From the above description, we know that different methods have their advantages, and the disadvantages are also obvious. Michael Rubinstein [9] defines the resizing space as a conceptual multi-dimensional space, combining several retargeting operators, including scaling, cropping, warping, seam carving, etc. They point that using several operators can potentially get better results for retargeting than using a single operator, and give an algorithm for finding an optimal multi-operator retargeting sequence under some assumptions.

Meanwhile, Achanta and Susstrunk [10] revise the definition of the energy function, and get an efficient, noise robust retargeting method based on seam carving by using saliency maps to assign higher importance to visually prominent whole regions (not just edges). There are many references contributed to obtaining saliency maps [11,12]. Itti model [1] is the most classical method that generates the saliency map by choosing orientation, intensity and color information as features, and combining these features' maps into a global saliency map. Besides, there are some other models proposed in recent years, including GAFFE [13] (Gaze-Attentive Fixation Finding Engine) model, the saliency model based on frequency-tuned detection [14,15] and the model based on inverse Fourier transform and phase spectrum [16]. Furthermore, some high-level features such as face-detection [14,17], are also taken into account to generate the more accurate saliency map.

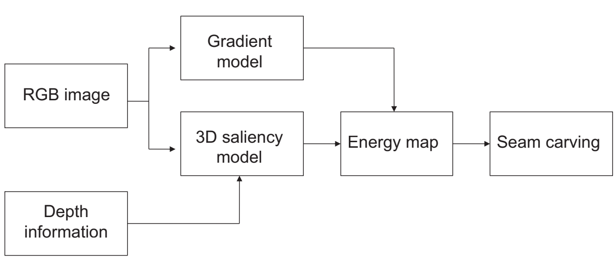

Fig. 1. Flow chart of our method. We obtain 3D saliency map from 2D features in RGB image and the depth information captured by 3D camera. Then we combine 3D saliency with gradient information coming from RGB image to guide seam carving algorithm.

With the development of technology, 3D information is accessible, i.e., kinect developed by Microsoft [28]. 3D information can help computer to understand the structure of image, which may be fail only by 2D features. Thus depth information [18–20] in 3D images is also taken into consideration in image retargeting recently. Shen [21] has applied the depth information into seam carving, and developed an efficient JND-based significant computation approach using the multi-scale graph cut based energy optimization. Lin [22] also notices the depth effect in seam carving, and obtains depth-based importance map automatically.

In this paper, we introduce saliency map into seam carving method. In other word, we use gradient information and saliency map together to determine the importance of each pixel. Moreover, we also use the depth information to update conventional 2D saliency into 3D saliency. Experimental results show that the proposed method does generate more desirable resized images than conventional seam carving algorithm.

2. Image retargeting using seam carving combining with 3D saliency

Seam carving is an image operator that supports content-aware image retargeting. The target of seam carving is to protect important information during resizing to make the processing unperceivable. The conventional seam carving method generally produce satisfactory results for only a class of images and it always suffering from distortion since the content of image has a variety of structure, and no single energy function performs well across all images. In this paper, we introduce a method which combine $L-1$ norm of gradient and 3D saliency to represent energy map. Due to the different characteristics of $L-1$ norm of gradient and 3D saliency in describing image, performance of seam carving using our method is more satisfactory than the state of the art approaches. Fig. 1 shows the flow chart of our method and more details are discussed in the following sections.

2.1. Seam carving with $L-1$ norm of gradient

Seam carving was first described in [8], as a content-aware algorithm for resizing image. It is an effective resizing approach by removing seams with less energy from image. Each seam is an 8-connected path of pixels on image from one side to the opposite. Suppose an $n \times m$ image $\boldsymbol{I}$ and a vertical seam can be defined as:

$$\boldsymbol{s^X} = \{s_i^X\} = \{(x(i), i)\}_{i=1}^{n}, \text{ s.t. } \forall i, |x(i) - x(i-1)| \le 1 \tag{1}$$

where x is a mapping $x: [1, \ldots, n] \rightarrow [1, \ldots, m]$, ensure each seam a 8-connected path of pixels, containing one, and only one pixel in each row. Similarly, if y is a mapping: $y: [1, \ldots, m] \rightarrow [1, \ldots, n]$, a

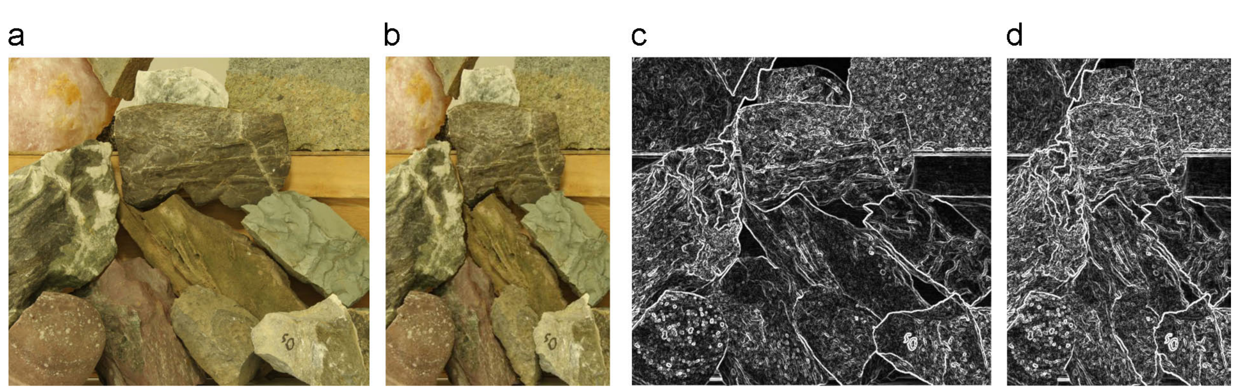

Fig. 2. $L-1$ norm of gradient performs well in seam carving. Due to the characteristic of gradient information, we can see that from (c) and (d), frame of image is preserved completely and the blanks between edges are removed in retargeting process.

horizontal seam is:

$$\mathbf{s}^{\mathbf{Y}} = \{s_i^Y\} = \{(y, y(i))\}_{i=1}^{n}, \text{ s.t.} \forall i, |y(i) - y(i-1)| \le 1 \quad (2)$$

Thus, given an energy function $\boldsymbol{E}$ to calculate the energy of each pixel, energy cost of each seam can be obtained. The seam $\boldsymbol{s}^*$ with minimum energy has priority to be removed, can be defined:

$$\boldsymbol{s}^* = \min_{\boldsymbol{s}} \boldsymbol{E}(\boldsymbol{s}) = \min_{S} \sum_{i=1}^{n} \boldsymbol{E}(s(i)) \quad (3)$$

where $s(i)$ is the pixel on the seam path.

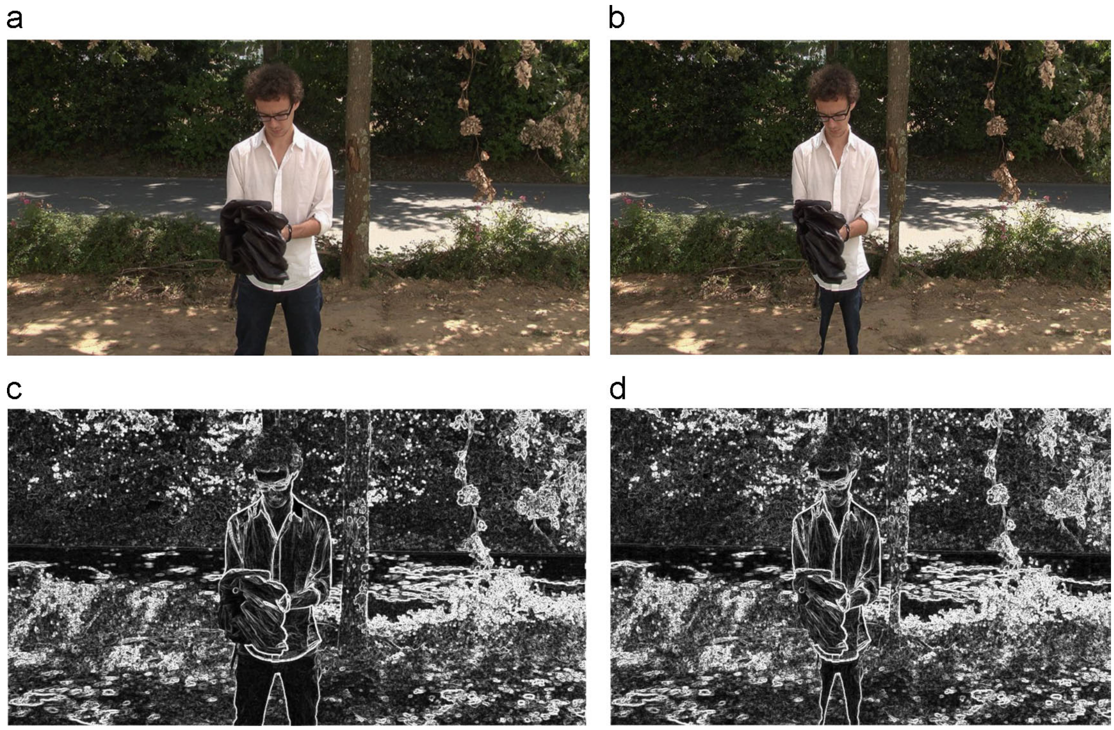

Fig. 3. Gradient features in seam carving may allow too many seams across object in some case, which may cause serious distortion by a low degree of retargeting. (b) Indicates this kind of distortion by resizing (a) to 90%. From (c) we can find the region of man's leg has a low gradient value that makes too many pixels in this part are removed in retargeting, as shown in (d).

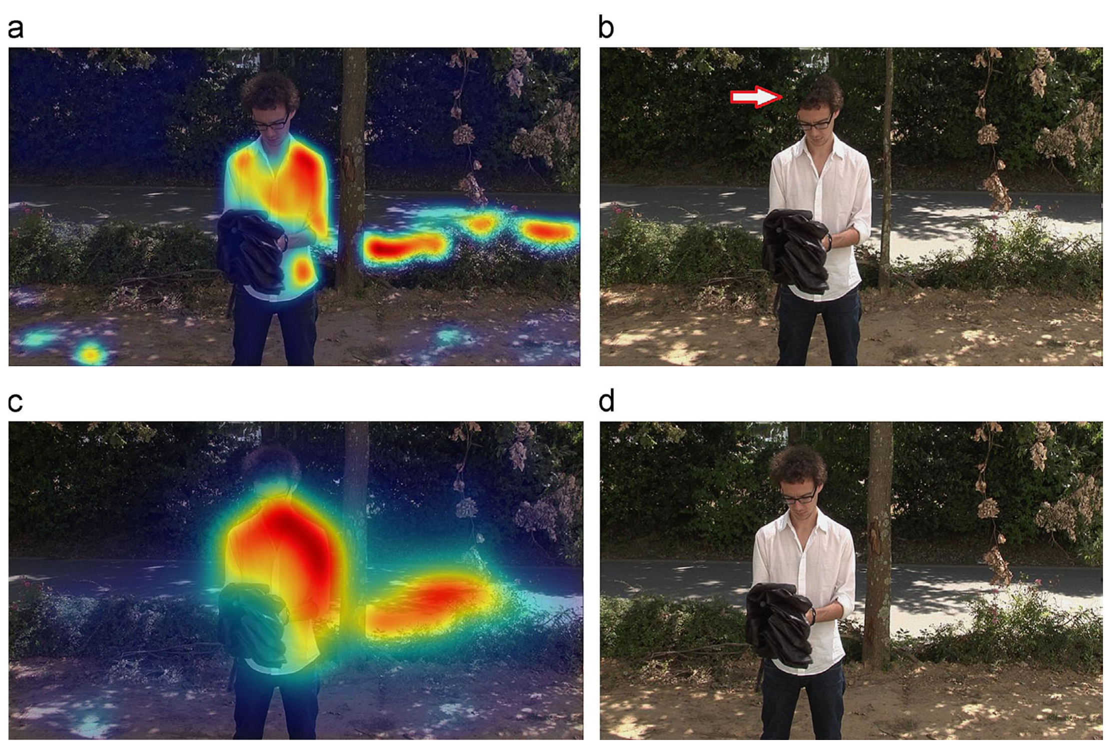

Fig. 4. From (a) and (c), GBVS seems more suitable for seam carving since it can detect object more accurately. By resizing original image to 85%, advantage of GBVS is shown in (d) compared to the result from FT shown in (b). Distortion can be perceived in (b) while (d) looks perfect.

It is obviously that energy function has a great impact on the performance of seam carving. Several visual features have been used as the energy function to guide seam carving and the most classic one is $L-1$ norm of gradient:

$$\boldsymbol{E}(x,y)=\left|\frac{\partial}{\partial x}\boldsymbol{I}(x,y)\right|+\left|\frac{\partial}{\partial y}\boldsymbol{I}(x,y)\right| \tag{4}$$

Gradient highlights edges in an image, in other words, it draws the outline of the frame in an image, which is very appropriate in seam carving. This characteristic of $L-1$ norm of gradient protects the frame of the whole image for seam removing and it has been proved to be effective on most images. Fig. 2 shows a satisfactory sample and indicates that the frame of image is still in good condition (Fig. 2(d)) after the width of original image resizing to 60%.

However, gradient cannot perform well all the time. Since gradient only concentrates on the edges of object, seams probably go across important object frequently in some cases, which may lead to an unacceptable deformation of salient object. Here is an example in Fig. 3, seam carving produces serious distortion while it only reducing 10% of original image in horizontal direction. It is easy to notice the problem by comparing gradient map in Fig. 3(c) and (d). The region of the man's leg has a low gradient value so as to allow too many seams across it, though it should be an important area to be protected.

2.2. Seam carving combining with 3D saliency

To overcome the shortcoming of gradient feature in seam carving illustrated in Fig. 3, we attempt to adopt 3D saliency which is more sensitive to object. Human visual system employs a saliency mechanism to limit processing to important information which is currently relevant to behaviors or visual tasks. This visual mechanism provides away to predict where users will interest in a

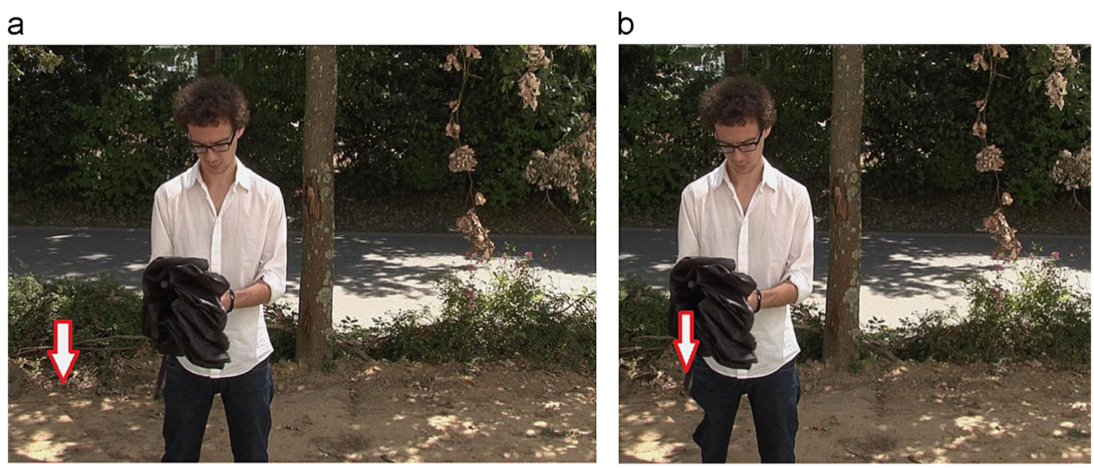

Fig. 5. With an increase degree of resizing, distortion caused by overprotection becomes serious.

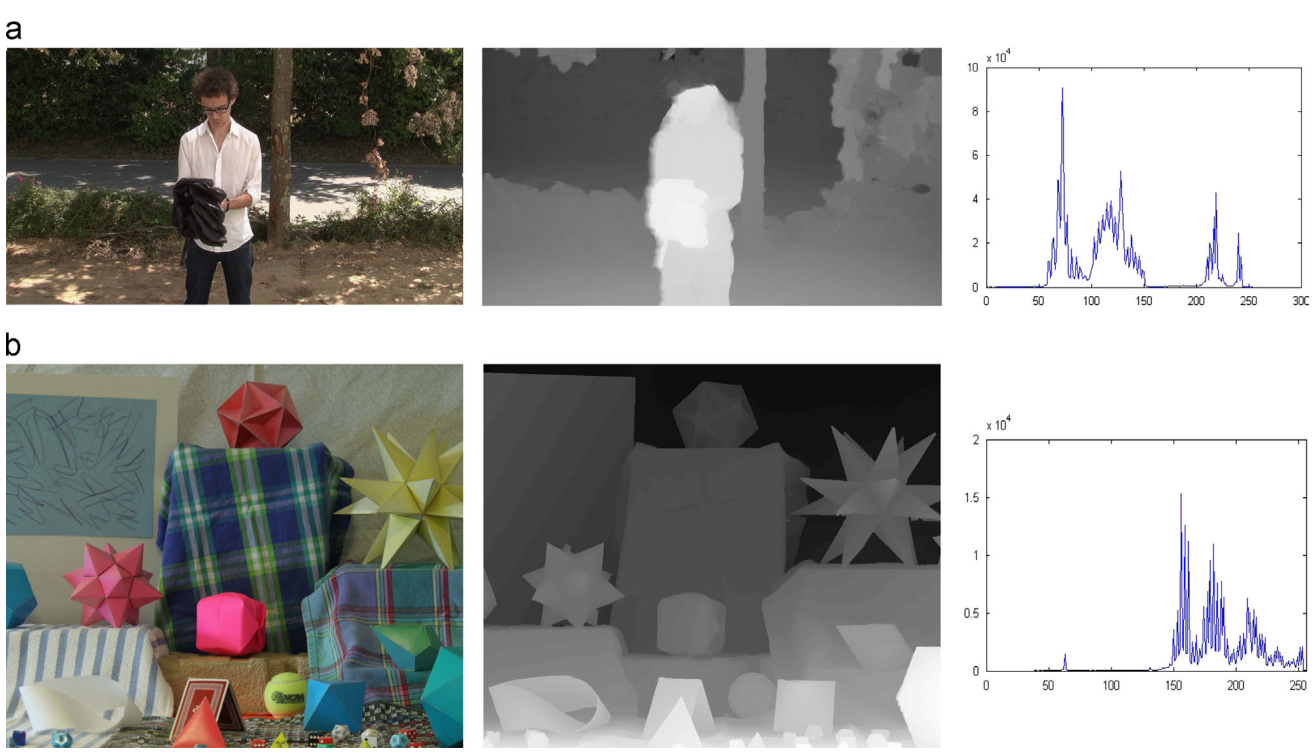

Fig. 6. Different structures of images have different dispersion of gray-scale of pixels. (a) and (b) show two types of images respectively with a salient object in it or not. From left to right, the original image, depth information and distribution of gray-scale of pixels corresponding to (a) and (b) are given.

scene, and can be used for seam carving because of the ability to preserve important object effectively.

2.2.1. Analysis of saliency-based seam carving

To quantitatively measure the conspicuity of a location, or the likelihood of a location to attract the attention of human observers, different models of saliency have been proposed. Each model differs in its selection of fixation points. Most models of saliency are biologically inspired and based on a bottom-up computational model. Typically, saliency is based on center-surround contrast of units modeled on known properties of primary visual cortical cells. First multiple low-level visual features such as intensity, color, orientation, and texture are extracted from the image at multiple scales. After a saliency map is computed for each of the features, they are normalized and combined in a linear or non-linear fashion into a master saliency map that represents the saliency of each pixel.

In Fig. 4, two commonly used saliency models of frequency-tuned salient region detection (FT) [14] and Graph-based visual saliency (GBVS) [23] are examined. We can find that better result comes from GBVS, which consists of two steps: first forming activation maps on certain feature channels, and then normalizing them in a way which highlights conspicuity and admits combination with other maps. The activation and normalization schemes in GBVS seem to be more suitable for images with salient objects and cluttered scenes.

By using GBVS instead of $L-1$ norm of gradient we achieve a better result (Fig. 4(d)) on the same image which has serious distortion shown in Fig. 3. Comparing (a) and (c), we can find the region of the man is protected more effectively by GBVS model than by FT model. In Fig. 4(b), serious distortion can be perceived on the left side of the man, especially upper-left of the man's head.

However, saliency model keeps salient object well at the expense of the rest part. All the saliency models in seam carving cannot avoid having a risk of 'frame broken'. Because of overprotect saliency parts, the frame of image may be broken when resizing too much or the salient object is dominant in scale. This kind of distortion becomes obvious with higher degree of resizing, as showed in Fig. 5.

2.2.2. Combination of gradient and 3D saliency

To prevent 'frame broken' happening and also protect salient object at same time, it is naturally to consider combing gradient feature and saliency map. In [24], they use weighted sum to combine gradient information, face saliency and body saliency together. Since saliency is a global relative value, the value of saliency in salient region is overwhelming compared to gradient value. It will be hard to choose a suitable weight coefficient to ensure the influence of gradient if we use weighted sum combination. In this paper, we use inner product to combine gradient information and 3D saliency map:

$$\boldsymbol{E}(x,y)=\boldsymbol{E_g}(x,y)\times\boldsymbol{E_s}(x,y) \tag{5}$$

where $\boldsymbol{E_g}(x,y)$ is gradient map and $\boldsymbol{E_s}(x,y)$ is 3D saliency map.

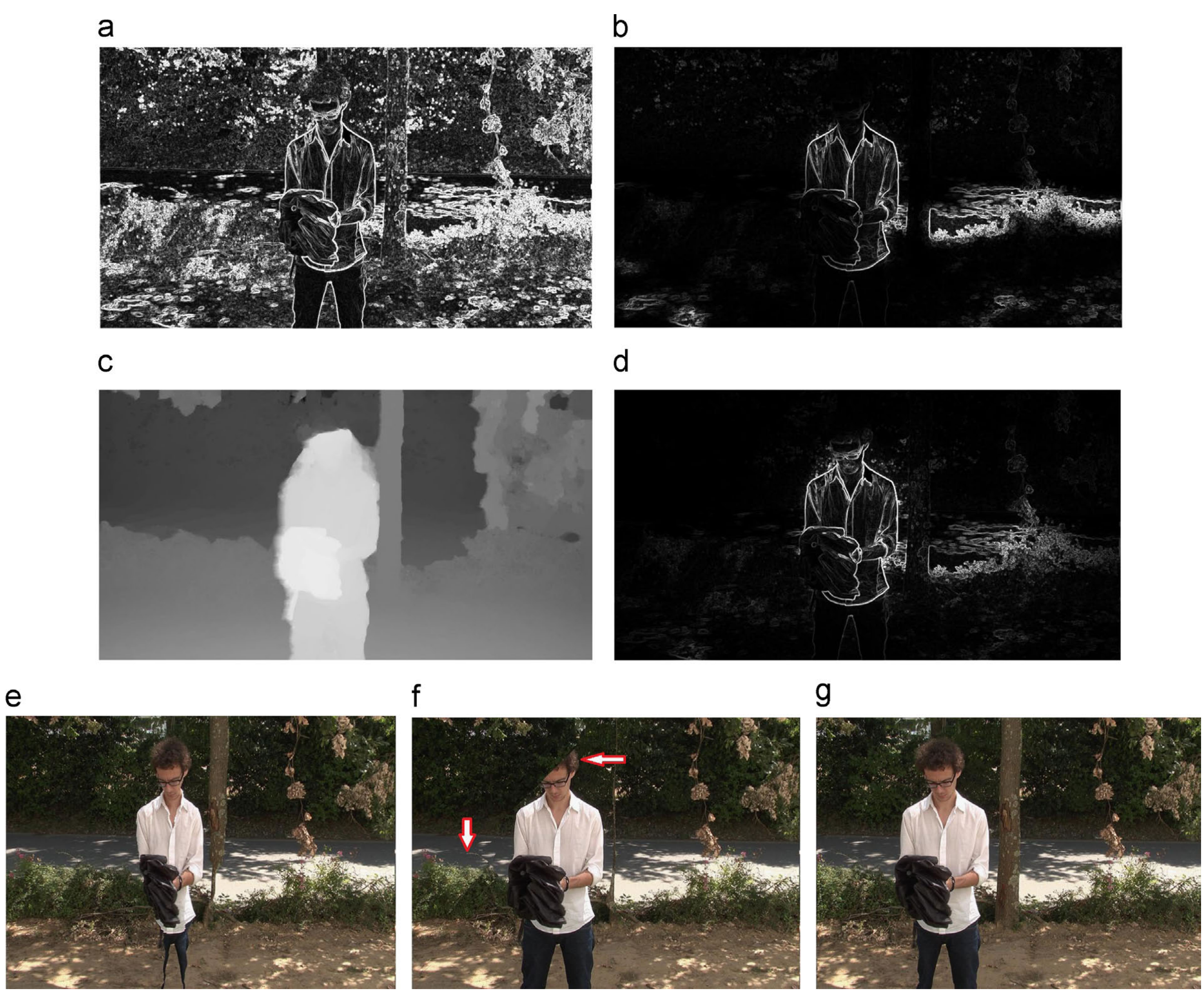

Fig. 7. (a) Original gradient map. (b) Gradient combined with 2D saliency. (c) Smooth depth map. (d) Gradient combined with 3D saliency. (e), (f), (g) are results by using energy function in [8,10] and our method.

Compared to traditional saliency model, 3D saliency has distinct advantage in seam carving application. Because the stereoscopic contents can provide additional depth cues that can be used in image understanding. Depth information helps saliency model to update 2D saliency map that comes from 2D features, such as color and intensity, which may limit or fail when the image has a complex background. Depth also provides a new viewpoint to analyze the structure of contents in the image, which will be propitious to seam carving. On the basis of the common sense that object near the observer will achieve more attention and more details will be perceived while tiny change happened. We define 3D saliency as:

$$\boldsymbol{E_s} = (1-\alpha)\boldsymbol{E_{2D}} + \alpha \times \boldsymbol{E_{depth}} \tag{6}$$

where $\boldsymbol{E_{2D}}$ is the GBVS saliency map, $\boldsymbol{E_{depth}}$ is the depth map captured by 3D camera and α is an adaptive coefficient.

By analyzing the gray-scale distribution of depth map, we know the dispersion of pixels in depth map, then we can tell which model is more fit to this image and adjust a. As shown in Fig. 6, image that has salient object or something dominate is more dispersed in gray-scale distribution (Fig. 6(a)), which has been proved to prefer saliency model. On the contrary, image which has a complex content and has non-salient object is more compact and stable in gray-scale distribution (Fig. 6(b)), and gradient information will be a better choice. To measure the dispersion D of pixels in image, we first quantize the gray-scale into K levels in the range of [0, 255], then use variance (Eq. (7)) to calculate dispersion. Note that we ignore the levels which have precious few pixels in it to avoid noise and idle calculation on minutia.

$$D = \sum_{i=1}^{K} n(i) \times \left(F_i - \overline{F}\right)^2 \tag{7}$$

where F_i is the ith quantitative value, $n(i)$ is the number of pixels belonging to the ith quantification intervals. We already know that an image which is half black and half white has the maximum value of D. Thus D_{max} is a constant value and we assume that a is linearly proportional to D. Based on this, we calculate D by Eq. 7, and use the ratio of D and D_{max} to adjust a in an adjustable interval.

Once achieved energy map, we use algorithm proposed by Grundmann in [25] to find optimal piecewise seam. The Discontinuous Seam-Carving algorithm proposed a spatial coherence

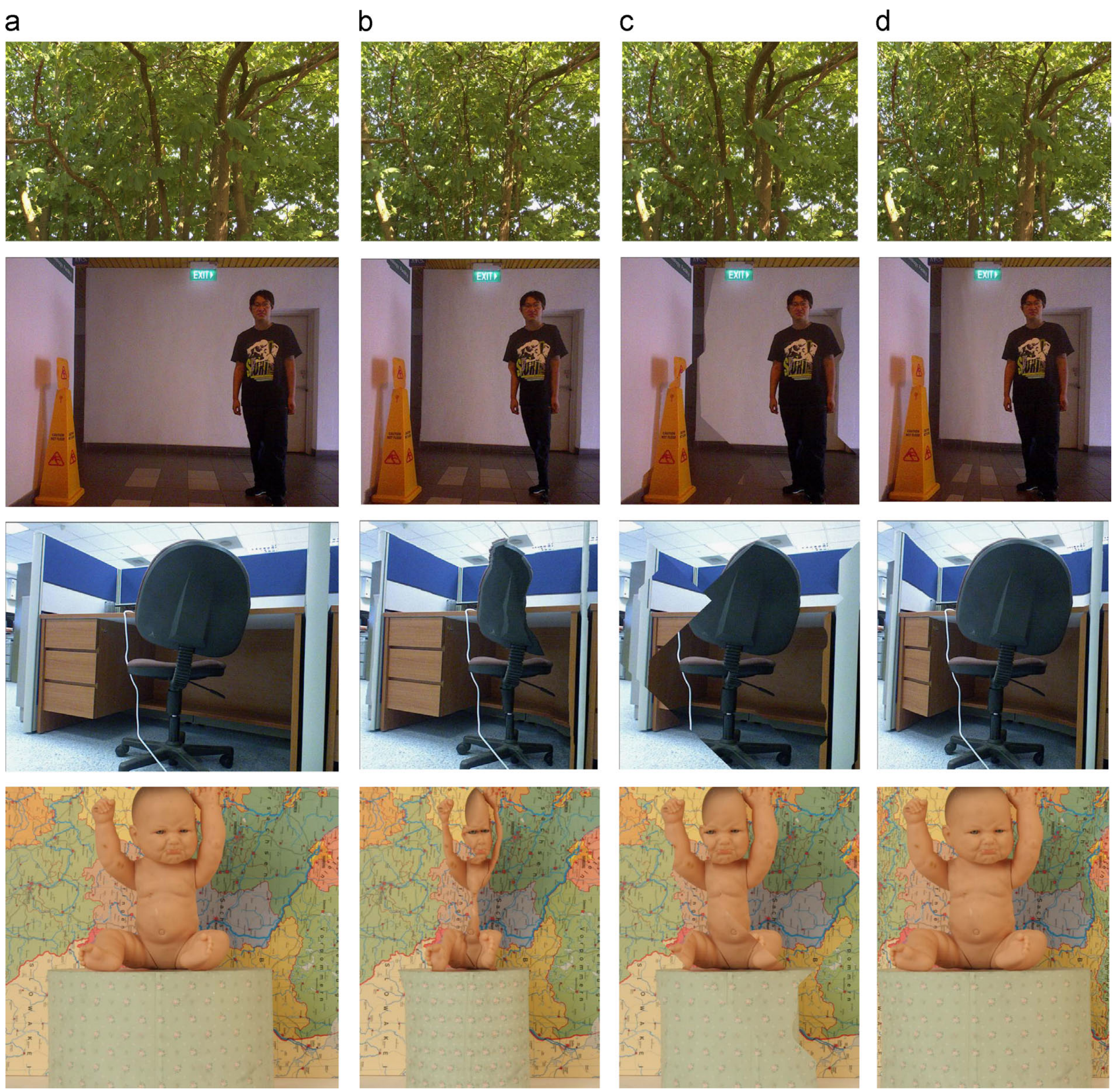

Fig. 8. Comparison between seam carving methods using different energy function. (a) The original image. (b) Result by using gradient as energy function. (c) Result by using GBVS saliency model only. (d) Our result.

measure which consider about the energy created by removing or inserting pixels during retargeting process. It breaks the rule that seam should be a 8-connected path and allows pixels to search not just three neighbors in the row above but some pixels within a limited window in that row. For a pixel (x_b, y) in the bottom row, the summed spatial transition cost to pixel $(x_a, y-1)$ in the top row $(x_a < x_b)$ is:

$$S_v(x_b, x_a, y) = \sum_{k=x_a}^{x_b-1} \left| G^v_{k,y} - G^d_{k,y} \right| + \sum_{k=x_{a+1}}^{x_b} \left| G^v_{k,y} - G^d_{k-1,y} \right| \tag{8}$$

where $G^v_{k,y} = |F_{k,y} - F_{k,y-1}|$ is the vertical gradient magnitude between pixel (k, y) and its top neighbor, while $G^d_{k,y} = |F_{k,y} - F_{k+1,y-1}|$ is its diagonal gradient magnitude with the top right neighbor. In practice, the optimal neighbor pixel is chosen in a window of 15 pixels around (x_b, y) to reduce computational cost.

Fig. 7 shows the improvement in seam carving by using our method. Frame and salient object are both preserved and the real objects achieve higher value in energy map while the rest is restrained. Process of our method is independent of parameters and large scale of experiments in next section demonstrates the effectiveness and superiority of our approach.

Fig. 9. Using algorithm in [8] (upper side) and our method (underside) to enlarge both (a) and (b). Our result is more satisfactory since salient object in human vision is preserved and structure of content in image is also complete.

3. Experiments and results

We implemented our seam carving method on 3DGaze database [26] and NUS3D-Saliency Dataset [27]. The stereoscopic images in the 3DGaze database are natural content images, acquired from two sources: (1) the Middlebury 2005/2006 image dataset, and (2) the IVC 3D image dataset. Each image in 3DGaze has the corresponding depth maps and eye-tracker data. The database in [27] includes indoor and outdoor scenes that have natural co-occurrence of common objects. They use Kinect camera [28], which consists of an infrared projector-camera pair as the depth camera that measures per pixel disparity to capture a 640×480 pixel color image and the corresponding depth image at the same time.

We compared our method with other existing approaches, showed in Fig. 8. In practice, we set α in the range of 0.2–0.8, considering the basic impact of both 2D saliency and depth information. In second column, gradient information in [8] can only work well when there is no salient object. In third column, overprotecting on salient region causes serious deformation by [10] when non-salient region is complex. Results of our method in fourth column are more satisfactory for different type of images. Local gradient information, global saliency information and stereoscopic depth information are all taken into consideration to make our method robust. Furthermore, we apply our method on extending process and it also performs well as showed in Fig. 9.

4. Conclusion

We proposed a novel method that use gradient information and 3D saliency model to guide seam carving. In our method, gradient information protect frame while saliency map protect objects, and depth information update energy map with 3D information making our result more comfortable. The combination is independent of parameters and fast. Experimental results have proved a great advantage of our method in application.

In future work, we will make effort to dig out more information from depth. By big data analyzing, a more effective and accurate 3D model can be set up to help image understanding. Due to the limitation of hardware equipment, improve the pretreatment on initial data coming from 3D camera is also necessary. Another interesting thing is to extend our framework into video retargeting process. Since the advantages in real-time performance of our method, we believe that a feasible approach for real-time application will be achieved.

Acknowledgements

This work is partially supported by National Nature Science Foundation of China (61105076), China Post-doctoral Science Foundation (2012M511402), Fundamental Research Funds for the Central Universities of China (2012HGCX0001).

References

[1] Laurent Itti, Christof Koch, Ernst Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[2] Laurent Itti, Christof Koch, Feature combination strategies for saliency-based visual attention systems, J. Electron. Imaging 10 (1) (2001) 161–169.
[3] Hasan, Muhammad Abul, Changick Kim. An automatic image browsing technique for small display users, in: ICACT 2009. Eleventh International Conference on Advanced Communication Technology, 2009. vol. 3. IEEE, 2009.
[4] Y.u.n. Liang, et al., Optimised image retargeting using aesthetic-based cropping and scaling, IET Image Proc. 7 (1) (2013) 61–69.
[5] Vidya Setlur, et al., Retargeting images and video for preserving information saliency, IEEE Comput. Graphics Appl. 27 (5) (2007) 80–88.
[6] Gal, R.a.n., Olga Sorkine, Daniel Cohen-Or. Feature-aware texturing. in: Proceedings of the 17th Eurographics Conference on Rendering Techniques. Eurographics Association, 2006.
[7] Wolf, Lior, Moshe Guttmann, Daniel Cohen-Or. Non-homogeneous content-driven video-retargeting, in: ICCV 2007. IEEE 11th International Conference on Computer Vision, 2007. IEEE, 2007.
[8] Shai Avidan, Ariel Shamir, Seam carving for content-aware image resizing (ACM), ACM Trans. Graphics (TOG), 26, 2007.
[9] Michael Rubinstein, Ariel Shamir, Shai Avidan, Multi-operator media retargeting (ACM), ACM Trans. Graphics (TOG) 28 (3) (2009).
[10] Achanta, Radhakrishna, Sabine Susstrunk. Saliency detection for content-aware image resizing, in: Sixteenth IEEE International Conference on Image Processing (ICIP), 2009. IEEE, 2009.
[11] Meng Wang, et al., Movie2comics: towards a lively video content presentation, IEEE Trans. Multimedia 14 (3) (2012) 858–870.
[12] Meng Wang, et al., Event driven web video summarization by tag localization and key-shot identification, IEEE Trans. Multimedia 14 (4) (2012) 975–985.
[13] Umesh Rajashekar, et al., GAFFE: a gaze-attentive fixation finding engine, IEEE Trans. Image Process. 17 (4) (2008) 564–573.
[14] Achanta, Radhakrishna, et al. Frequency-tuned salient region detection, in: CVPR 2009. IEEE Conference on Computer Vision and Pattern Recognition, 2009. IEEE, 2009.
[15] Schauerte, Boris, Rainer Stiefelhagen. Predicting human gaze using quaternion dct image signature saliency and face detection, in: IEEE Workshop on Applications of Computer Vision (WACV), 2012. IEEE, 2012.
[16] Guo, Chenlei, Q.i. Ma, Liming Zhang. Spatio-Temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE, 2008.
[17] Moran Cerf, et al., Predicting human gaze using low-level saliency combined with face detection, NIPS (2007).
[18] Elezovikj, Semir, Haibin Ling, Xiufang Chen. Foreground and scene structure preserved visual privacy protection using depth information, in: IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2013. IEEE, 2013.
[19] Seung-Won Jung, K.o. Sung-Jea, Depth sensation enhancement using the just noticeable depth difference, IEEE Trans. Image Process. 21 (8) (2012) 3624–3637.
[20] Dahan, Meir Johnathan, et al., Combining color and depth for enhanced image segmentation and retargeting, Visual Comput. 28 (12) (2012) 1181–1193.
[21] Shen, Jianbing, Dapeng Wang, Xuelong L.i. Depth-Aware Image Seam Carving. (2013): 1-1.
[22] Lin, Wei-Yang, Pei-Chen Wu, Bo-Rong Chen. Image Retargeting Using Depth Enhanced Saliency. Proceedings of 3DSA 2013 7: 1.
[23] Jonathan Harel, Christof Koch, Pietro Perona, Graph-based visual saliency, Adv. Neural Inf. Process. Syst. 19 (2007) 545.
[24] Hwang, Daw-Sen, Shao-Yi Chien. Content-aware image resizing using perceptual seam carving with human attention model, in: IEEE International Conference on Multimedia and Expo, 2008. IEEE, 2008.
[25] Grundmann, Matthias, et al. Discontinuous seam-carving for video retargeting, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010.
[26] Junle Wang, et al. (1-1), A Computational Model of Stereoscopic 3D Visual Saliency (2013).
[27] Congyan Lang, et al., Depth Matters: Influence of Depth Cues on Visual Saliency. Computer Vision–ECCV 2012, Springer, Berlin Heidelberg (2012) 101–115.
[28] Kinect: ⟨http://www.xobx.com/kinect⟩.

Yanxiang Chen received the Ph.D. degree from Electronic Science and Technology in University of Science & Technology of China (USTC) in 2004. As a visiting scholar, she has been to Beckman Institute at University of Illinois at Urbana-Champaign (UIUC) from August 2006 to April 2008, and Department of Electronic Computer Engineering at the National University of Singapore (NUS) from November 2012 to February 2013. She is a SPIE member, ACM member, and CCF member. Her work has been supported by the Natural Science Foundation of China, Natural Science Foundation of Anhui Province of China, and Postdoctoral Science Foundation of China. Her current research fields include audio-visual signal processing, saliency and scene analysis.

Yifei Pan received the B.S. degree in Electrical & Information Engineering from Hefei University of Technology, Hefei, Anhui, China, in 2011. He is currently pursuing the master degree in multimedia at Hefei University of Technology. His current research interests include saliency detection in image and object tracking in video.

Minglong Song received the B.S. degree in Communication Engineering from Hefei University of Technology, Hefei, Anhui, China, in 2013. He is currently pursuing the master degree in Hefei University of Technology. His main research interests include image retargeting, saliency and video retargeting.

Meng Wang received the B.E. and Ph.D. degrees in the Special Class for the Gifted Young and the Department of Electronic Engineering and Information Science from the University of Science and Technology of China (USTC), Hefei, China, respectively.

He is a Professor in the Hefei University of Technology, Hefei, China. He previously worked as an Associate Researcher at Microsoft Research Asia, and then a core member in a startup in Silicon Valley. After that, he worked in the National University of Singapore as a Senior Research Fellow. He has authored more than 100 book chapters, journal, and conference papers in these areas. His current research interests include multimedia content analysis, search, mining, recommendation, and large-scale computing.

Dr. Wang is a member of ACM. He was the recipient of the Best Paper Awards continuously in the 17th and 18th ACM International Conference on Multimedia and the Best Paper Award in the 16th International Multimedia Modeling Conference.