AUTOMATIC IMAGE RETARGETING EVALUATION BASED ON USER PERCEPTION

Tongwei Ren and Gangshan Wu

State Key Laboratory for Novel Software Technology, Nanjing University, China

ABSTRACT

As image retargeting techniques have attracted more and more attention for effective image display on mobile devices, quality evaluation of image retargeting is required. To address the lack of automatic evaluation techniques in retargeting, this paper proposes a user perception based framework to automatically evaluate the quality of the target image against the original image. In the framework, the pixels in the original image and the target image are first approximately order-preserved matched by dynamic programming. Based on the pixel matching result, several features are extracted to describe the user requirements and further adapted to fit user perception in retargeting. Finally, the overall score of the target image quality is calculated by integrating the scores in different evaluation aspects. Experiments demonstrate the effectiveness of the proposed framework.

Index Terms**—** Image retargeting, automatic evaluation, user perception, pixel matching

1. INTRODUCTION

The prevalence of mobile devices, such as cellular phones and PDAs, lead to the significant need of convenient image display on the screens with small sizes and arbitrary aspect ratios [1]. However, the existing images usually have much higher resolutions and their aspect ratios are nearly fixed. To overcome the mismatch, image retargeting techniques are proposed to adapt the original image to a concise target image which fits the target screen requirement. A common solution of image retargeting is uniform scaling, but it may cause important details loss or distortion. So the current image retargeting methods focus on content-aware sampling based on pixel energy [2][3] or image similarity [4][5].

Different to the prosperity in image retargeting techniques research, the target image quality evaluation has attracted little attention. The lack of an effective evaluation mechanism leads to the problem that most existing image retargeting methods have to demonstrate their performance through several selected examples of comparison with other methods, which are insufficient and easy to cause misleading in evaluation. Some methods use manual evaluation on image dataset instead of few examples [3], which obtains the evaluation results close to user perception but has high labor cost and time cost. Moreover, manual evaluation is not suitable for large scale evaluation, for it is hard to keep the consistent criteria among many evaluators or during a long evaluation period.

Due to the limitation of manual evaluation, automatic evaluation techniques for image retargeting are highly demanded [6]. Unfortunately, as far as we know, such technique is still unavailable. Some optimization based image retargeting methods propose their objectives in retargeting, such as bidirectional similarity (BDS) [4] and bi-directional warping (BDW) [5], which can be treated as the analogues of automatic evaluation criteria. However, these optimization objectives only consider the coverage of the original image content but ignore the user perception. They can't provide a comprehensive consideration of the target image quality and their evaluation results may distinctive to user perception.

To address the problems of current work in image retargeting evaluation, we propose a novel automatic quality evaluation framework providing human like evaluation for image retargeting. Considering a user viewing images prefers to obtain the entire information for precise understanding and insensible artifacts for good experience, an effective target image should emphasize the important content while retaining sufficient context information with little visual distortions. Therefore, we establish our evaluation criteria of the target image quality including important content emphasis (CE), global information coverage (IC) and visual distortion reduction (DR). With the evaluation criteria, we constitute our framework by combining automatic image quality evaluation. Current automatic image quality evaluation methods [6][7] are roughly classified into three categories, full-reference, reduced-reference and no-reference [6]. In our framework, we focus on reduce-reference quality evaluation, which can obtain higher accuracy than no-reference evaluation and keep practicable when the full-reference method is unavailable. Fig. 1 shows the overview of our proposed framework. We first construct an approximate order-preserved pixel matching between the original image and the target image, which is treated as the reference in our evaluation. Then, we extract several features to describe the corresponding evaluation criteria. Based on the features, we calculate the scores in different criteria and integrate the scores to provide an overall evaluation of the target image quality. To the best of our knowledge, it is the first automatic evaluation framework for image retargeting which can provide user perception based evaluation.

2. ORDER-PRESERVED PIXEL MATCHING

Generally speaking, image retargeting is a procedure of discrete or continuous pixel sampling based on energy map(s) or image similarity. The change of pixel locations is the key factor influencing the quality of the target image. Hence, in our

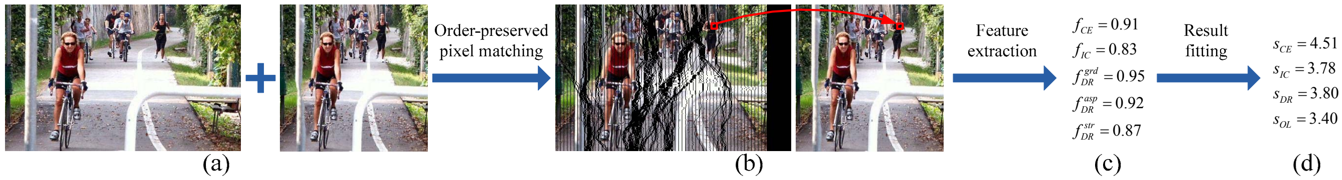

Fig. 1. An overview of the proposed framework. (a) Original image and target image. (b) Result of approximately order-preserved pixel matching. (c) Features to describe user perception based evaluation criteria. (c) Automatic evaluation scores.

framework, we match the pixels between the original image and the target image, and evaluate the retargeting performance by observing the location change of each pixel.

A meaningful pixel matching between the original image and the target image should preserve the pixel orders. However, the order-preserved pixel matching problem has been proved to be NP-complete [8]. Considering the characters of image retargeting, we utilize dynamic programming to approximately match pixels in polynomial time.

We first consider the situation that the retargeting only happens in one direction, such as horizontal direction. For a given original image I_S of size $W_S \times H_S$ and the corresponding target image I_T of size $W_T \times H_T$, we use dynamic programming to match their pixels in the *kth* rows:

$$M[i,j] = \max\{M[i-1,j-1] + \varphi(I_S[i,k], I_T[j,k]), \\ M[i-1,j],\ M[i,j-1]\}, \tag{1}$$

where $I_S[i,k]$ and $I_T[j,k]$ are the *ith* and *jth* pixels in the *kth* rows of the original image and the target image respectively; $\varphi(.)$ equals one when the distance between two pixels in Lab color space is smaller than the pre-defined threshold (using 0.2 in our experiments) and zero otherwise. The algorithm can well solve the pixel matching between the original image and the target image in one dimension, and it is $O(W_S W_T H_S)$ in time.

The pixel matching between the original image and the target image generated by two-dimension retargeting will be more complex, for the pixels in the same row/column of the target image may be from the different rows/columns of the original image. Here, we utilize the strategy similar to the asymmetric-DTW algorithm in [5]. We first match each row of the target image with the rows in the original image based on Equation (1). For the row orders are not inversed in image retargeting, the *kth* row of the target image is only required to match the rows of the original image in the range of $[k, k+h]$, here $h = H_S - H_T$. In this way, we can obtain a row similarity matrix between the original image and the target image and further calculate the best row matching by dynamic programming, which provides a rough matching result including most pixels in the target image. To each unmatched pixel $I_T[i,j]$, we search its matched pixel in the coordinate range of $[i, i+w] \times [j, j+h]$ by assuming the pixel orders are preserved in retargeting, here $w = W_S - W_T$. In this way, we can generate the approximately order-preserved pixel matching between the original image and target image, and the total time cost is $O(hW_S W_T H_S)$.

3. USER PERCEPTION BASED EVALUATION

According to the evaluation criteria, we evaluate the target image quality in the aspects of important content emphasis, global information coverage and visual distortion reduction.

3.1. Important Content Emphasis

Emphasizing the important content is the basic purpose of image retargeting. A direct method is to segment the original image and calculate the area ratio of the important objects in the target image. However, the accurate object segmentation is still an open problem, and the either-or judgment of "important objects" is also very hard in many situations. Hence, instead of calculating the area ratio of important objects, we assign an importance value to each pixel and evaluate the ratio of retained energy in retargeting:

$$f_{CE} = \frac{\sum_{i=1}^{W_S}\sum_{j=1}^{H_S}\left(E[i,j]\cdot\Phi[i,j]\right)}{\chi_{W_T H_T}\left(E[i,j]\right)}, \tag{2}$$

where $E[i,j]$ denotes the importance energy of $I_S[i,j]$, which can be automatically detected [9] or manually defined; $\Phi[i,j]$ equals one if $I_S[i,j]$ has a matched pixel in the target image and otherwise it equals zero; $\chi_K(S)$ is the sum of the K largest elements in the set *S*.

3.2. Global Information Coverage

With the size limitation, the target image can only retain partial pixels of the original image. However, the information of the discarded pixels may be covered by their nearby retained pixels. We define the information coverage between two pixels as follows:

$$cog\,(I[i,j], I[x,y]) = sim(I[i,j], I[x,y]) \cdot \exp\left(-\frac{(i-x)^2+(j-y)^2}{2}\right), \tag{3}$$

where $sim(.)$ is the similarity of two pixels in Lab color space, $\exp(.)$ denotes the location distance between two pixels.

Then, we calculate the whole information coverage of the target image to the original image. To each pixel in the original image, we find the pixel whose matched pixel in the target image can maximally represent its information:

$$f_{IC} = \frac{\sum_{i=1}^{W_S}\sum_{j=1}^{H_S}\max_{I_S[x,y]}\left(cog(I_S[i,j], I_T[x',y'])\cdot\Phi[x,y]\right)}{W_S H_S}, \tag{4}$$

where $I_T[x',y']$ is the matched pixel of $I_S[x,y]$.

3.3. Visual Distortion Reduction

Visual distortion is the common artifact degrading the user experience in image retargeting. It is usually caused by three reasons: the generated new edges, the changed aspect ratio of important content, and the distorted structure information. In the following, we deal with each aspect separately.

The new edges are generated by the pixel location changes in sampling, so we calculate the sum of the gradients between the pixels are nonadjacent in the original image:

$$f_{DR}^{grd} = 1 - \frac{\sum_{i=1}^{W_T}\sum_{j=1}^{H_T}\left(\left|\frac{\partial}{\partial i} I_T[i,j]\right|\cdot\delta_i[i,j] + \left|\frac{\partial}{\partial j} I_T[i,j]\right|\cdot\delta_j[i,j]\right)}{2W_T H_T}, \quad (5)$$

where $\left|\frac{\partial}{\partial i} I_T[i,j]\right|$ and $\left|\frac{\partial}{\partial j} I_T[i,j]\right|$ are the gradient of $I_T[i,j]$ in horizontal direction and vertical direction respectively; $\delta_i[i,j]$ equals zero when $I_T[i,j]$ and $I_T[i+1,j]$ are the adjacent pixels in the same row of the original image, and otherwise it equals one; $\delta_j[i,j]$ is defined similarly.

To evaluate the aspect ratio change of important content, we decompose the original image into the patches in size of $n\times n$ (using 16×16 in our experiments) and calculate the aspect ratio change of each patch in retargeting. For a patch after retargeting may not be rectangular, we simply define its aspect ratio as the quotient of the sum of its top and bottom boundaries' length len_{TB} to the sum of its left and right boundaries' length len_{LR}. It is calculated as follows:

$$f_{DR}^{asp} = 1 - \frac{1}{K}\sum_{k=1}^{K}\left(E(k)\cdot\left(\max\left(\frac{len_{LR}(k)}{len_{TB}(k)}, \frac{len_{TB}(k)}{len_{LR}(k)}\right) - 1\right)\right), \quad (6)$$

where K is the number of patches; $E(k)$ is the average pixel importance energy of the *kth* patch.

The general structure information in image is hard to describe. Here, we only consider the straight lines, which can be detected by Hough transformation or manual interaction. To each line l_k in the original image, we find the matched pixels of its pixels in the target image and generate a line l'_k with the least sum of squared distances to all these pixels. Then, we evaluate the structure information distortion as follows:

$$f_{DR}^{str} = 1 - \frac{1}{K'}\sum_{k=1}^{K'} d_{\max}(k), \quad (7)$$

where K' is the number of lines; $d_{\max}(k)$ is the maximal distance of the corresponding matched pixels to l'_k.

3.4. Fitting to User Perception

The extracted features are highly related to the user perception, but they may be not similar in value. To provide human like evaluation results, we fit the features to manual evaluation results by utilizing linear regression in each evaluation aspect. In DR score calculation, we use the minimum value among f_{DR}^{grd}, f_{DR}^{asp} and f_{DR}^{str}, for the users are usually care about the worst element in evaluation:

$$D_k \begin{pmatrix} s_{CE} \\ s_{IC} \\ s_{DR} \end{pmatrix} = \begin{pmatrix} \alpha_{CE} \\ \alpha_{IC} \\ \alpha_{DR} \end{pmatrix} + \begin{pmatrix} \beta_{CE} & 0 & 0 \\ 0 & \beta_{IC} & 0 \\ 0 & 0 & \beta_{DR} \end{pmatrix} \begin{pmatrix} f_{CE} \\ f_{IC} \\ \min(f_{DR}^{grd}, f_{DR}^{asp}, f_{DR}^{str}) \end{pmatrix},$$

where $s_{\#}$ are the automatic evaluation scores in *CE*, *IC*, and *DR* respectively.

Considering the users prefer to give an overall evaluation (OL) of the target image quality, we further integrate the scores of all aspects by linear combination:

$$s_{OL} = \alpha_{OL} + \beta_{OL}^{CE} s_{CE} + \beta_{OL}^{IC} s_{IC} + \beta_{OL}^{DR} s_{DR}. \quad (9)$$

4. EXPERIMENTS

We collect 100 images with different contents and styles by search engine, and generate the corresponding target images using six typical image retargeting methods, including uniform scaling, cropping [2], seam carving [1], non-homogenous warping [9], constrained region warping [3], and multi-operator method [5]. Then, we invite six evaluators in age of 22 to 37, including students, officers and company employees, to carry out the manual evaluation using five levels mean opinion score (MOS). Each target image is randomly distributed to three evaluators and evaluated in the aspects of CE, IC, DR and OL respectively. To eliminate the personal evaluation preference, the manual evaluation results are preprocessed as $s'_{\#,k} = s_{\#,k} - \overline{s}_{\#,k} + \overline{s}_{\#}$, here $s_{\#,k}$ is a manual evaluation result by the *kth* evaluator in the aspect #, $\overline{s}_{\#,k}$ is the mean value of his/her evaluation results to all target images in this aspect, and $\overline{s}_{\#}$ is the mean value of all manual evaluation results in this aspect. We randomly select 50 original images, using their corresponding target images and manual evaluation results as the training data, and treat the rest as the test data.

We first extract the features and fit them to the manual evaluation results in the training data. Tab.1 shows the parameters of fitting. Then, we compare the automatic evaluation scores to the manual evaluation results in the test data. Tab. 2 shows a comparison example of the images in Fig. 2. We can find the automatic evaluation results are similar to manual evaluation results and in accord with intuitive perception, e.g. result #1 and #6 have serious distortion and result #2 and #3 are weak in information coverage. Fig. 3(a)-(d) shows the comparison between MOS and automatic evaluation results in the whole test data in the aspects of CE, IC, DR, and OL respectively, and the correlation coefficient (*CC*) and mean absolute error (*MAE*) are shown in the last two rows of Tab. 1. It shows the automatic evaluation results generated by the proposed framework well match the user perception.

We also compare the proposed framework with the existing optimization objectives used in image retargeting, such as BDS [4] and BDW [5]. From Fig. 3(d)-(f), we can find the results of our proposed framework match the manual evaluation results better. The essential reason is that these objectives only focus on image similarity but ignore the user perception, for example, they don't consider the difference of image content importance in user perception.

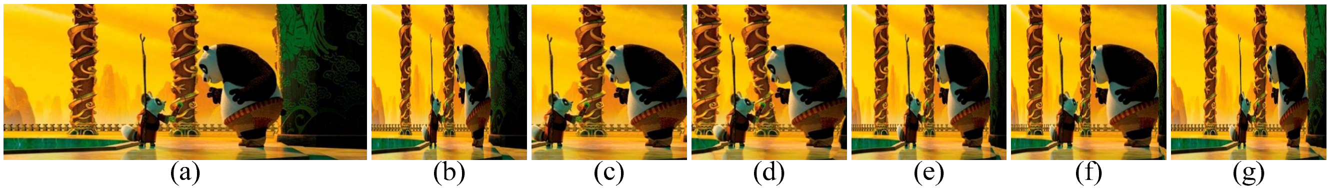

Fig. 2. Example of the original image and the target images. (a) Original image. (b)-(g) Target images generated by different image retargeting methods, orderly numbered from #1 to #6.

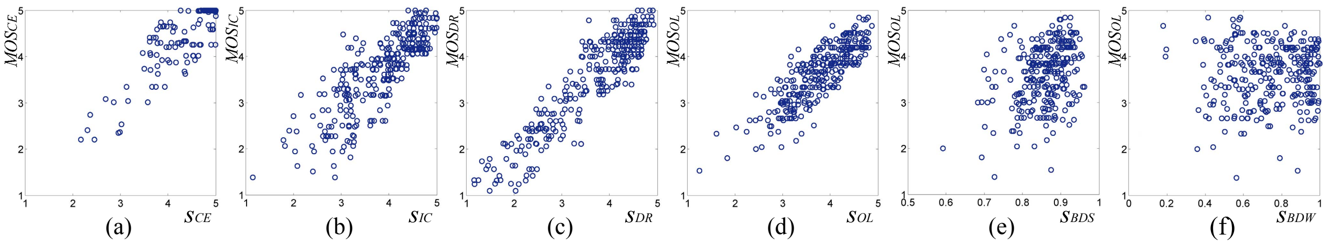

Fig. 3. Scatter plots of mean opinion score (MOS) versus automatic evaluation results. (a) CE. (b) IC. (c) DR. (d) OL. (e) MOS in OL versus BDS. (f) MOS in OL versus BDW.

Tab. 1. Parameters and results of automatic evaluation

	CE	IC	DR	OL
α	-0.81	-1.34	-1.61	-4.74
β	5.85	6.17	6.22	0.67, 0.66, 0.69
CC	0.90	0.82	0.92	0.85
MAE	0.24	0.38	0.22	0.28

Tab. 2. Example of comparison between MOS and automatic evaluation scores

	MOS				Automatic Score			
	CE	IC	DR	OL	CE	IC	DR	OL
#1	3.69	4.82	2.62	2.04	3.48	4.71	2.75	2.60
#2	5.00	2.03	4.83	3.42	4.95	2.37	4.94	3.55
#3	4.82	2.54	3.69	3.71	4.94	3.02	3.93	3.28
#4	4.24	4.51	4.72	4.33	4.39	4.58	4.43	4.28
#5	4.73	4.53	4.60	4.51	4.67	4.49	4.65	4.56
#6	3.89	4.42	3.11	3.00	3.87	4.56	3.02	2.95

In experiment, we also find some limitations of our approach. One difficult problem is the trade-off on dataset size. The larger dataset provides more training data and test data but it requires more evaluators or longer evaluation time, which will introduce more inconsistence in evaluation and degrade the performance of our framework. Another limitation is some features are only approximately implemented for automatically extraction, such as structure information distortion.

5. CONCLUSION

This paper proposes an automatic quality evaluation framework of image retargeting for the first time. The evaluation criteria derived from real user requirements in retargeting are proposed. With the criteria, three underlying algorithms, including approximate order-preserved pixel matching, user perception based feature extraction, and automatic evaluation score fitting, are utilized in the framework. Together, they provide a comprehensive evaluation of the target image quality.

Our future work will focus on improving the manual evaluation and feature descriptors in our framework. We will also consider the possibility to apply the proposed criteria in image retargeting techniques and extend the framework to video retargeting evaluation.

6. ACKNOWLEDGEMENTS

This work is supported by Grant 60721002, 60975043 of the National Natural Science Foundation of China and the Hong Kong General Research Fund PolyU 5204/09E.

7. REFERENCES

[1] S. Avidan and A. Shamir, "Seam Carving for Content-Aware Image Resizing," *ACM Transactions on Graphics*, 26(3):1-9, 2007.
[2] L.Q. Chen, X. Xie, X. Fan, W.Y. Ma, H.J. Zhang, and H.Q. Zhou, "A Visual Attention Model for Adapting Images on Small Displays," *Multimedia System*, 9(4):353-364, 2003.
[3] T. Ren, Y. Liu, and G. Wu, "Image Retargeting Using Multi-Map Constrained Region Warping," *ACM Multimedia*, China, 2009.
[4] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing Visual Data Using Bidirectional Similarity," *IEEE International Conference on Computer Vision and Pattern Recognition*, USA, 2008.
[5] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator Media Retargeting," *ACM Transactions on Graphics*, 28(3):1-11, 2009.
[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *Transactions on Image Processing*, 13(4):600-612, 2004.
[7] T. Ren, Y. Liu, and G. Wu, "Full-Reference Quality Assessment for Video Summary," *Workshop on Video Mining*, Italy, 2008.
[8] D. Keysers, and W. Unger, "Elastic Image Matching is NP-Complete," *Pattern Recognition Letter*, 24(1-3): 445-453, 2003.
[9] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous Content-driven Video-retargeting," *IEEE International Conference on Computer Vision*, Brazil, 2007.