Image Retargeting Assessment based on Salient Region Similarity

Jun Wang[1], Tongwei Ren[1,3,*], Yan Liu[2], Jia Bei[1,3], Bowei Xu[1]
[1]Software Institute, Nanjing University, 210093, Nanjing, China
[2]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[3]State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
xnervwang@gmail.com, csyliu@comp.polyu.edu.hk, {rentw, beijia, xbw09}@software.nju.edu.cn

ABSTRACT

The prosperity of image retargeting technique leads to the significant need of effective approaches for assessing different image retargeting methods. In this paper, we propose a novel automatic approach providing human perception based image retargeting assessment by measuring salient region similarity between the original image and the target image. First, the salient regions in the original image and the target image are matched using gradient, color and direction features. Then, the similarity between the salient regions are measured based on feature distances and region saliency. Based on salient region similarity, the quality of target image is assessed with two criteria derived from real user requirements in retargeting, important content retainment and visual artifact reduction, and the overall scores of the target image quality are finally calculated by integrating assessment results on different criteria. Experiments demonstrate the effectiveness of the proposed approach.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

Algorithms, Measurement, Human Factors

Keywords

Image retargeting, quality assessment, human perception, salient region similarity

1. INTRODUCTION

The prevalence of mobile devices calls for personalized image display on the screens with small sizes and arbitrary aspect ratios [3]. To avoid important content loss or distortion in resizing, numerous content-aware image retargeting methods are proposed. These retargeting methods change the sizes and aspect ratios of the original images by shrinking and/or expanding the less important content and generate the target images with retained important content and insensible visual artifacts.

Though all the image retargeting methods try to optimize their effects, a study of the existing image retargeting methods shows no method can ensure the generated target image satisfactory for any original image [11]. It is becoming important to assess the target image quality to help improving the performance of the corresponding retargeting method. Furthermore, since some retargeting method may work best on certain original images while another method on the different ones, recommending the best target image by assessing different target images will improve the user experience in viewing target images.

Manual image retargeting evaluation is one possibility to obtain assessment results close to human perception. Rubinstein et al. did a comparative study of existing image retargeting methods [11] and provided a public dataset "RetargetMe" as a benchmark for further image retargeting assessment research [1]. Castillo et al. used eye-tracking data to evaluate different image retargeting methods [4]. For high labor cost and time cost, manual evaluation is not suitable for large scale evaluation or applications requiring rapid response. Due to the limitations of manual evaluation, automatic assessment for image retargeting is highly demanded. Some optimization objectives used in image retargeting can be treated as the analogues of automatic assessment criteria, but they have been found in low agreement with human perception [11]. Ren et al. approximately computed the quality factors extracted from user perception [10], but it depends on the effectiveness and efficiency of pixel matching. Liu et al. used top-down manner to organize global and local features in assessment [7], but it is found that the users prefer sacrificing content over inserting deformation to the target images [11].

To address these problems, we propose a novel automatic assessment approach providing human like quality evaluation for image retargeting. Current automatic image quality assessment methods can be roughly classified into three categories, full-reference, reduced-reference and no-reference [12]. For it is too difficult to obtain a perfect target image as the reference or assess the target image in a totally blind manner, we assume the original image with perfect quality and focus on reduced-reference assessment. In our approach, we pay attention to two criteria, important content retainment (ICR) and visual artifact reduction (VAR), and assess the quality of target image based on salient region similarity

ICIMCS '12, September 9-12, 2012, Wuhan, China

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

Figure 1: An overview of the proposed approach. (a) Original image. (b) Target image. (c) Salient region matching result, here the number of matched regions is reduced for viewing. (d) Automatic assessment results in important content retainment and visual artifact reduction. (e) The overall scores for reference evaluation and no-reference evaluation.

(SRS) between the original image and the target image. Fig. 1 shows an overview of our approach. We first detect the salient regions in the original image and the target image. Then, we construct a salient region matching and treat it as the reference in assessment. Based on salient region similarity, the target image is assessed important content retainment and visual artifact reduction, and the overall scores of target image quality are further calculated for different evaluation situations.

2. SALIENT REGION SIMILARITY BASED ASSESSMENT

2.1 Criteria

While viewing a image, a user usually prefers to obtain the entire information for precise understanding and insensible artifacts for good experience. Hence, an effective target image should retain the important content and cause little visual artifacts in image resizing. We establish our assessment criteria of the target image quality including important content retainment and visual artifact reduction.

In general, important content retainment and visual artifact reduction require high level image understanding, which are hard to directly measure. However, in image retargeting, the target image is generated from the original image by shrinking/expanding the less important content. By comparing the original image and the target image, we can find how many salient regions in the target image can be matched to the original image content. We treat the matched salient regions as the retained important content from the original image and the unmatched regions as the generated visual artifacts. In this way, the quality assessment of target image can be formulated to a problem of matching the salient regions between the original image and the target image and measure their similarities.

2.2 Salient Region Detection and Matching

There are many salient region detection methods for image matching, which detect the interest points from the images and extract the feathers from the local regions around the points. SIFT [8] is a widely used local features, and it has been used in image retargeting assessment [5]. However, SIFT has several limitations for matching the original image and the target image. To enhance robustness in matching, SIFT ignores color change and region rotation in matching, which may increase the mismatching and weak visual distortion detection. And SIFT uses Euler distance to measure the similarity between two regions, which cannot handle the tiny distortions with few influence to image viewing.

We improve the salient region detection and matching method for special requirements in image retargeting assessment. We first detect the interest points and determine the region scales and dominant directions in a similar way to SIFT [8]. For obvious color change is seldom happened in image targeting, we extract both gradient histogram and HSV color histogram from each region as the feature, which can well reduce the mismatching between the regions. Then, we match the regions in the original image and the target image based on gradient histogram, HSV color histogram and the angle between two regions' dominant directions. Here, two matched regions should satisfy all the following requirements: the earth movers' distance of gradient histogram satisfies the matching strategy in [9], the Euler distance of HSV color histograms and the angle between region dominant directions are less than the predefined thresholds (0.2 and 15° in our experiments respectively).

2.3 Similarity based Assessment

To the matched salient regions r_i and r_j, we calculate their similarity as follows:

$$sim(r_i, r_j) = sim(f_i^{gra}, f_j^{gra}) \cdot sim(f_i^{HSV}, f_j^{HSV}) \cdot \cos\theta. \quad (1)$$

where $sim(f_i^{gra}, f_j^{gra})$ denotes the similarity of gradient histogram, $sim(f_i^{HSV}, f_j^{HSV})$ denotes the similarity of HSV color histogram, θ is the angle between the dominant directions of r_i and r_j.

Based on the salient region similarity, the quality of the target image is assessed with two criteria: important content retainment and visual artifact reduction. Important content retainment is assessed by measuring the proportion of the matched salient regions in the original image:

$$score_{ICR} = \frac{\sum_i s_i^{ori} M(r_i^{ori})}{\sum_i s_i^{ori}}. \quad (2)$$

where $M(r_i^{ori})$ denotes whether a region r_i^{ori} in the original image has a matched region $r_{i'}^{tar}$ in the target image, which equals $sim(r_i^{ori}, r_{i'}^{tar})$ if r_i^{ori} has a matched region and 0 otherwise. s_i^{ori} is the saliency value of r_i^{ori}. We use a low cost method [2] to calculate the saliency of each pixel in the original image, and the mean saliency value of all the pixels in a region is treated as the saliency value of the region.

Visual artifact reduction is assessed by measuring the proportion of the matched salient regions in the target image:

$$score_{VAR} = \frac{\sum_j s_j^{tar} M(r_j^{tar})}{\sum_j s_j^{tar}}. \quad (3)$$

where $M(r_j^{tar})$ denotes whether a region r_j^{tar} in the target image has a matched region $r_{j'}^{ori}$ in the original image, which equals $sim(r_j^{tar}, r_{j'}^{ori})$ if r_j^{ori} has a matched region and 0 otherwise. s_j^{tar} is the saliency value of r_j^{tar}, which is calculated as above.

Based on Equation (2) and (3), we calculate the overall score of the target image quality by integrating the assessment results in ICR and VAR:

$$score = w_{ICR} \cdot score_{ICR} + w_{VAR} \cdot score_{VAR}. \tag{4}$$

where *score* is the final assessment result of the target image quality. w_{ICR} and w_{VAR} are the nonnegative weights of ICR and VAR respectively, and $w_{ICR}+w_{VAR}=1$.

3. EXPERIMENTS

3.1 Dataset

We verify the proposed approach on RetargetMe dataset [1]. RetargetMe dataset contains 37 original images with one or more attributes, such as lines/edges, symmetry and so on. And each original image has 8 corresponding targeted images generated by nonhomogeneous warping (WARP), seam carving (SC), scale-and-stretch (SNS), multi-operator (MULTIOP), shift-maps (SM), streaming video (SV), uniform scaling (SCL) and manual cropping (CR), respectively [11].

RetargetMe dataset also provides two versions of manual evaluation results, reference version and no-reference version. The difference of the two versions of manual evaluation is the original image was shown in reference evaluation and not shown in no-reference evaluation. In each version, 210 participants are invited to vote the better target image in paired comparisons, and each target image can obtain up to 63 votes [11].

3.2 Weight Selection

In manual evaluation, the participants usually have different judgements when shown the original images as the references or not. So it requires different w_{ICR} and w_{VAR} values in the overall assessment score calculation for reference evaluation and no-reference evaluation.

To each original image, there are $\binom{8}{2}$= 28 target image pairs to compare. We randomly select 8 target image pairs for each original image, and calculate the overall assessment scores using different w_{ICR} and w_{VAR} with 0.1 per step. We compare the assessment scores and manual votes in each target image pair, and count the concordant target image pairs in reference evaluation and no-reference evaluation respectively. Here, "concordant" means the target image with higher assessment score has higher manual votes in a pair. Experiments show that the proposed approach has the most concordant target image pairs when $w_{ICR} = 0.6, w_{VAR} = 0.4$ in reference evaluation and when $w_{ICR} = 0.3, w_{VAR} = 0.7$ in no-reference evaluation, respectively. We use these w_{ICR} and w_{VAR} values in the following experiments.

The weights selection result shows important content retainment influenced the manual evaluation results even in no-reference evaluation, though the original images were not shown as the references. One possible reason is the participants could obtain a rough impression of the original image when view the target images, especially the results generated by uniform scaling, and they might pay attention to the content loss in evaluation.

3.3 Agreement with Manual Evaluation

3.3.1 Using Kendall Rank Correlation Coefficient

To measure the agreement between manual evaluation and automatic assessment results, we analysis the assessment results with Kendall correlation coefficient [6]:

$$\tau = \frac{\sum_{i=1}^{n_{ori}\binom{n_{tar}}{2}} C(i)}{n_{ori}\binom{n_{tar}}{2}}. \tag{5}$$

where $C(i)$ denotes the concordance of the ith image pair, which equals 1 if the target image with higher automatic assessment score has higher manual votes in the ith image pair and -1 otherwise. $n_{ori} = 37$ is the number of original images. $n_{tar} = 8$ is the number of target images to each original image.

Fig. 2 shows a comparison of the proposed approach with other automatic assessment methods in agreement with manual evaluation, including bidirectional similarity (BDS), bidirectional warping (BDW), color layout (CL), edge histogram (EH), Earth-Mover's Distance (EMD) and SIFT-flow (SIFTflow). All the automatic assessment results of the compared methods are provided by RetargetMe dataset [1]. As shown in Fig. 2, the proposed approach obtains 48.9% and 54.1% agreement with manual evaluation in reference evaluation and no-reference evaluation, which is much better than other methods.

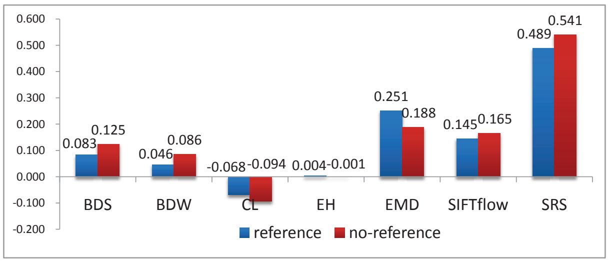

Figure 2: Comparison in agreement measurement using Kendall rank correlation coefficient.

3.3.2 Using Weighted Rank Correlation Coefficient

Though the proposed approach obtains better performance than other methods while using Kendall rank correlation coefficient, the agreement between the proposed approach and manual evaluation is still low. We analyze of the manual evaluation results, and find the votes are very similar in some image pairs and quiet distinctive in other image pairs. It means the participants have more ambiguous even inconsistent judgements in some image pairs than in others, in which the discordance of the assessment scores and the manual votes should be less penalized. In this regard, Kendall rank correlation coefficient is inappropriate. The difference of manual votes should be considered to treat the concordances and discordances between the manual votes and the assessment scores in various image pairs differently in agreement measurement. So we utilize weighted rank correlation coefficient to analysis the automatic assessment results:

$$\tau^* = \frac{\sum_{i=1}^{n_{ori}\binom{n_{tar}}{2}} w_i C(i)}{\sum_{i=1}^{n_{ori}\binom{n_{tar}}{2}} w_i}. \tag{6}$$

where $C(i)$ denotes the concordance of the ith image pair, which equals 1 if the target image with higher automatic assessment score has higher manual votes in the ith image pair and -1 otherwise. w_i is the weight of the ith image pair, which equals the vote difference of the two target images in the ith image pair.

Fig. 3 shows a comparison of the proposed approach with the same six automatic assessment methods in agreement with manual evaluation. It is found that most automatic assessment methods achieve better agreement with manual evaluation, and the proposed approach keeps ahead in both reference evaluation and no-reference evaluation by obtaining 65.4% and 71.2% agreement with manual evaluation respectively.

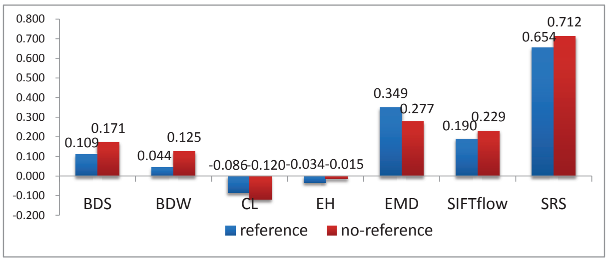

Figure 3: Comparison in agreement measurement using weighted rank correlation coefficient.

3.4 Discussion

In experiment, we also find some limitations of our approach. For example, while the dominant object occupies large area in the original image (Fig. 4(a)), our approach will be ambiguous in low distortion and better content retainment. In Fig. 4(b), the reference/no-reference assessment scores of the target images generated by CR, MULTIOP and WARP are 0.805/0.610, 0.790/0.623 and 0.803/0.650, while the manual votes are 42/46, 40/34 and 14/26 respectively.

Moreover, our approach cannot work well when the saliency detection is not accurate. In Fig. 4(c), Obama's body is similar to the background in color, which leads to low saliency and missing distortion detection. In Fig. 4(d), the reference/no-reference assessment scores of the target images generated by CR and SCL are 0.880/0.819 and 0.903/0.838, while the manual votes are 55/48 and 33/31 respectively.

4. CONCLUSION

In this paper, we propose an automatic assessment approach for image retargeting. Two assessment criteria derived from human perception in retargeting, including important content retainment and visual artifact reduction. With these criteria, image retargeting assessment is formulated to a salient region similarity measurement problem between the original image and the target image. Salient regions are matched by gradient, color and direction features, and the similarity between salient regions are measured based on feature distances and region saliency. Based on salient region similarity, the quality of target image is assessed on two criteria, and they provide the overall scores of the target image quality together.

Our future work will focus on improving image retargeting techniques with the proposed assessment approach. We will also consider the possibility to extend the approach to video retargeting assessment.

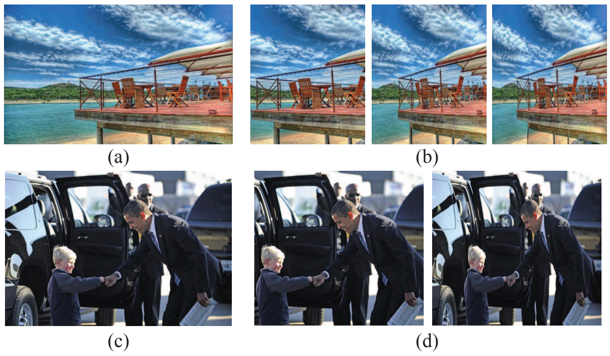

Figure 4: Examples of bad results. (a) Original image (Deck). (b) Target images generated by CR, MULTIOP and WARP. (c) Original image (Obama). (d) Target images generated by CR and SCL.

5. ACKNOWLEDGMENTS

This paper is supported by Natural Science Foundation of Jiangsu Province (BK2012304), Fundamental Research Funds for the Central Universities (1117021705, 1118021717), the Research Fund of State Key Laboratory for Novel Software Technology at Nanjing University (ZZKT2011B04), and the National Undergraduate Innovation Program (201210284031).

6. REFERENCES

[1] Retargetme dataset. *http://people.csail.mit.edu/mrub/retargetme/*.

[2] R. Achanta, S. Hemamiz, F. Estraday, and S. Süsstrunk. Frequency-tuned salient region detection. *IEEE CVPR*, 2009.

[3] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM TOG*, 26(3):10, 2007.

[4] S. Castillo, T. Judd, and D. Gutierrez. Using eye-tracking to assess different image retargeting methods. *ACM SIGGRAPH APGV*, 2011.

[5] S. Hua, G. Chen, H. Wei, and Q. Jiang. Similarity measure for image resizing using sift feature. *EURASIP*, 6:1–11, 2012.

[6] M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[7] Y. Liu, X. Luo, Y. Xuan, W. Chen, and X. Fu. Image retargeting quality assessment. *Computer Graphics Forum*, 30(2):583–592, 2011.

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[9] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. *ECCV*, 2008.

[10] T. Ren and G. Wu. Automatic image retargeting evaluation based on user perception. *IEEE ICIP*, 2010.

[11] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM TOG*, 29(6):160, 2010.

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.