

REGISTRATION BASED RETARGETED IMAGE QUALITY ASSESSMENT

Bo Zhang[†], Pedro V. Sander[†], Amine Bermak,^{†*} *IEEE FELLOW*

[†]Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

* Hamad Bin Khalifa University, Education City, Doha, Qatar

ABSTRACT

In recent years, a large number of image retargeting methods have been proposed. Measuring their relative quality is of significant importance, and there is still room for improvement in the effectiveness of objective retargeted image quality assessment (RIQA) metrics. In this paper, we propose a registration based RIQA metric. First, we propose to calculate the flow map using an image registration method which involves SURF point matching and halfway domain optimization. Using the computed flow map and the source image, we propose an LGI metric which contains three factors: 1) local similarity which assesses the local aspect ratio change, edge directional similarity and flow smoothness; 2) global distortion which measures the appearance change of salient objects; 3) salient information loss. Comparing with other six metrics, our LGI metric correlates the best with subjective rankings on the RetargetMe dataset.

Index Terms— retargeted image quality assessment (RIQA), dense correspondence, salient object segmentation, edge direction similarity

1. INTRODUCTION

With the development of display devices, there is a growing need for content-aware image retargeting methods in order to adapt the image display on the screen of different aspect ratios [1, 2]. The image retargeting methods can be roughly categorized into two classes. Discrete methods, such as cropping (CR), seam-carving (SC) [3], shift-map (SM) [4] and MULTIOP [5], directly remove the pixels or patches under the guidance of visual importance map. On the other hand, the continuous methods, such as uniform scaling in one dimension (SCL), non-homogeneous warping (WARP) [6], streaming video (SV) [7] and scale-and-stretch (SNS) [8], aim to generate a continuous pixel-wise flow that morphs the source image to the target size.

The retargeted image quality assessment (RIQA) is quite different from the traditional image quality assessment tasks, because the image to be assessed is not aligned to the reference, and the characteristic of the human visual system (HVS) to geometric distortion is far from being fully understood. In [1], Rubinstein et al. conducted a comprehensive study of dif-

ferent retargeted methods. Based on their dataset, they also examined the performance of several objective retargeted image quality metrics, including BDS [9], BDW [5], EH [10], SIFTflow [11] and EMD [12]. However, most of those metrics are generalized distance metrics, and do not correlate well with the subject rankings on assessing retargeted images. Many recent RIQA works use SIFTflow for image registration, and then measure the similarity between the corresponding patches. The structural similarity index (SSIM) is utilized in [13] to measure the structural similarity of the local patches, and the saliency map is used for pooling the local similarity values. In [2], the geometric distortion is measured based on the variation of the flow field. In [14], retargeted distortions are assessed by the change of the local aspect ratio. However, all of these works treat the image and the flow field separately. Also, the intrinsic geometric characteristics of the image, such as the local edge direction and strength, are not taken into consideration.

In this work, we use a halfway domain method for robust image registration. Then, using the flow mapping, a new RIQA metric called LGI is proposed. The main contributions of this work include: (1) a halfway domain registration method is proposed for robust image registration; (2) a novel local similarity is proposed by jointly considering the image content and the morphing flow; (3) an efficient salient object detection method is proposed to detect the salient objects for the global geometric distortion assessment.

2. ALGORITHM SCHEME

The whole framework of our LGI metric is illustrated in Fig. 1. We first propose a novel registration method that establishes the dense correspondence between the source and retargeted images on the halfway domain. The optimization method can find the dense correspondence robustly, even when the retargeted method introduces a large geometric distortion or content removal. After the optimization, a flow field that for pixel correspondence with sub-pixel accuracy is obtained.

In the second part of this work, we use the flow and the source image to assess the geometric similarity from three perspectives. First, local similarity value (LSV) uses the local aspect ratio change, the edge directional change and the

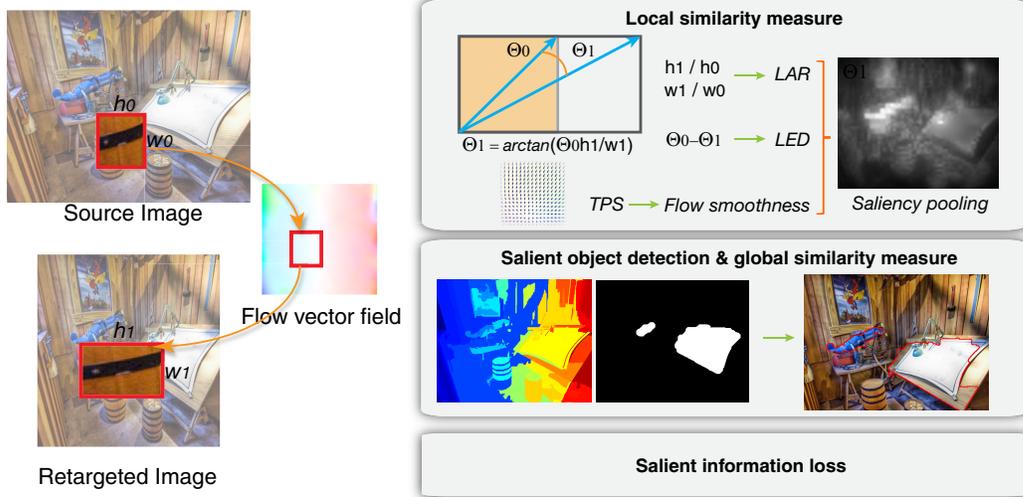


Fig. 1: The framework of the RIQA metric. The flow map that corresponds the patches of two images are first generated. Using the flow map and the source image, the local similarity, global similarity and information loss are computed.

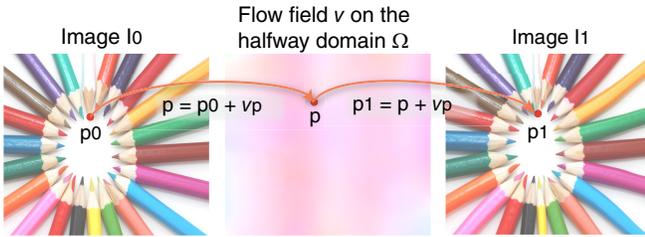


Fig. 2: The point p_0 is morphed through p to p_1 , where p is the corresponding point on the halfway domain Ω .

smoothness of the local flow as local distortion measure, and these local values are then pooled according to the saliency map. Second, global similarity value (GSV) is measured by appearance change of the objects which are segmented out using saliency map. Third, salient information loss is considered as the perceptual content loss.

3. DENSE CORRESPONDENCE

Inspired by the work in [15], we find the dense correspondence between the input images in a halfway domain. Suppose the image temporally morphs from the source image I_0 to the retargeted image I_1 (Fig. 2). On the halfway domain Ω , the morphing flow is defined by a 2D vector field v . The point p on the halfway and the corresponding points $p_0 \in I_0$, $p_1 \in I_1$ have such relationship:

$$p = p_0 + v_p \quad \text{and} \quad p_1 = p + v_p \quad (1)$$

Next, we will compute the flow on the halfway domain Ω for pixel correspondence. One advantage of optimizing on the

halfway is that, even in the presence of simple occlusion or point degeneration, the flow on Ω is still continuous, which makes the regularization simple [15]. Also, both the forward and backward flow, which are used for assessing geometric distortion and information loss respectively, can be obtained from single halfway optimization.

To compute the flow vector that best aligns the pixels between images, we optimize the energy function on the halfway domain:

$$E(v) = \sum_{p \in \Omega} E(p) \quad (2)$$

where $E(p) = E_{color}(p) + \alpha_1 E_{TPS}(p) + \alpha_2 E_{SURF}(p)$ (3)

The first term E_{color} is the chrominance distance in the CIELAB color space. The second term is the thin-plate spline (TPS) energy formulated as [16], which penalizes the smoothness of the flow vector.

For most of cases, the optimization using the first two terms can give a satisfactory result. However, for images containing textures and repetitive patterns, the chance of patch mismatch will increase. For better robustness, we introduce SURF point matching to constrain the solution of the flow, because the feature points appears more discriminant for sparse correspondence. In our work, we first find 100~200 pairs of SURF points matched by RANSAC [17]. Denote pairs of matched SURF points in I_0 and I_1 as u_i^0 and u_i^1 respectively, then the corresponding point $\hat{u}_i = (u_i^0 + u_i^1)$ on the halfway is supposed to have a flow vector of $v_{\hat{u}_i} = (u_i^1 - u_i^0)/2$, and this serves as a soft constraint for the energy function in Eq. (3):

$$E_{SURF}(p) = \|v_p - v_{\hat{u}_i}\|^2 \quad (4)$$

For the pixels $p \in \Omega$ that are not constrained by the SURF pairs, we set E_{SURF} to zero.

To solve the optimization efficiently, we adopt a coarse-to-fine approach. On each image scale, we employ the gradient descent method. After convergence, the point p_0 in I_0 can find its corresponding point p on the halfway through the iterative search [15]:

$$p^{(0)} = p_0 \quad (5)$$

$$p^{(i+1)} = p_0 + v_{p^{(0)}} \quad (6)$$

Then, the forward flow which corresponds the point p_0 to p_1 will be $v_{p_0} = 2v_p$. The backward flow can also be computed similarly. Using the backward flow, the contents that are preserved in I_1 during the retargeting will be marked, thus forming a mask map M which represents the areas of shared content.

4. PROPOSED RETARGETED IQA METRIC

In this section, we will describe the image quality assessment method using the source image along with the flow map generated from the halfway domain optimization.

4.1. Local Similarity

To assess the retargeted image quality, it is important to examine the local geometry similarity patch by patch between the images. In order to increase the computational efficiency, the source image I_0 is first divided into patches with size $w_0 \times h_0$. The corresponding patches in the retargeted image I_1 can be found using the forward flow. Because of the retargeting manipulation, the patch in I_1 may have a different size of $w_1 \times h_1$.

4.1.1. Local Aspect Ratio (LAR) Similarity

Ideally, a patch without any distortion keeps its aspect ratio and absolute size. The degree of the aspect ratio change influences the local similarity between images. To measure the degree of local aspect ratio change, the height and width change ratios are denoted as $R_h = h_1/h_0$ and $R_w = w_1/w_0$, and the patch sizes in I_0 and I_1 are denoted by $S_0 = w_0h_0$ and $S_1 = w_1h_1$ respectively. Then the local aspect ratio similarity (LAR) for the k_{th} patch is given by:

$$LAR_k = \left(\frac{2R_w R_h + C}{R_w^2 + R_h^2 + C} \right) \left(\frac{2S_0 S_1 + C}{S_0^2 + S_1^2 + C} \right) \quad (7)$$

where, C is a small constant. LAR_k ranges from (0, 1].

4.1.2. Local Edge Direction (LED) Similarity

The LAR values are calculated using the flow vector only. However, from Fig. 3 we find that, patches may also have different levels of geometric distortion: for the oblique edge, the aspect ratio change influences its edge direction greatly; on the other hand, the edge in the vertical direction is not

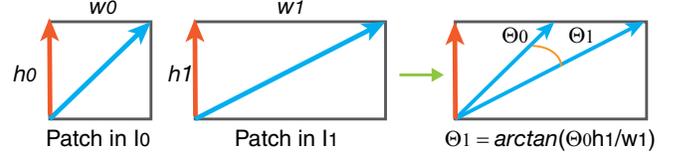


Fig. 3: The aspect ratio change only induces the directional change for the oblique blue edge.

affected. This enlightens us to design a metric by jointly considering the flow and the image edges. The edge of the source image is first located using the structured edge detection method from [18]. For the k_{th} patch, its principal edge direction Θ_0 is the weighted average of edge directions θ_i according to the edge strength e_i in the patch:

$$\Theta_0 = \sum_i \theta_i e_i, \quad \Theta_0 \in [-\pi/2, \pi/2] \quad (8)$$

After the retargeting, the patch is morphed to have an aspect ratio of w_1/h_1 , as shown in Fig 3. The principal edge direction will change to be:

$$\Theta_1 = \arctan\left(\frac{\tan(\Theta_0)h_1}{w_1}\right), \quad \Theta_1 \in [-\pi/2, \pi/2] \quad (9)$$

Then, the local edge direction (LED) similarity between Θ_0 and Θ_1 is defined as:

$$LED_k = e^{k_1(\Theta_0 - \Theta_1)^2 \cdot E} \quad (10)$$

Here $E = \sum_i e_i / (w_0 h_0)$ is the average edge strength, which is used to attenuate the drop of LED_k for the patches with small edge strength. For instance, when the edge strength is weak and $E = 0$, the edge direction change should not affect the visual quality, therefore the LED_k is 1.

4.1.3. Local Flow Smoothness (LFS)

In our work, we also propose to use TPS energy in Eq.(3) to evaluate the local flow smoothness (LFS):

$$LFS_k = e^{-k_2 E_{TPS}} \quad (11)$$

We use k_2 for normalization relative to other metrics.

4.1.4. Saliency Pooling

Finally, the perceptual local similarity value (LSV) is computed by combining the the above three terms:

$$LSV = \frac{\sum_k (LAR_k \cdot LED_k \cdot LFS_k)^{LCM_k} \cdot S_k \cdot M_k}{\sum_k S_k \cdot M_k} \quad (12)$$

where, S_k is the patch saliency calculated by the method in [19]; M_k is the mask map, representing the preserved patch

area in the retargeted image. In Eq. (12), LCM is the local confidence map ($LCM = [2\mu_0\mu_1 + C]/[\mu_0^2 + \mu_1^2 + C]$, μ_0 and μ_1 are the mean intensity of the corresponding patches), and represents the confidence of correct patch matching. If the registration is not accurate, the LCM for that patch will be close to 0 and the local similarity of that patch does not affect the LSV.

4.2. Global Similarity

In order to assess the geometric distortion for retargeted images, the assessment at the object level is also important, as it relates to the more semantic human perception of images. In order to calculate the global geometric similarity, we first segment the image with the statistical region merging method [20] (Fig. 1). We compute the saliency map using the DRFI model [21]. Then, connected homogeneous regions that have saliency values larger than a threshold are further merged. Up to three largest regions are finally chosen. Fig. 1 shows the result of salient object detection.

Suppose m salient objects are detected. The n th salient object in the source image has the size of $S_n^{(0)} = H_n^{(0)}W_n^{(0)}$. Using the forward flow, the object is morphed to have a size of $S_n^{(1)} = H_n^{(1)}W_n^{(1)}$. Denote $R_n^{(h)} = H_n^{(1)}/H_n^{(0)}$ and $R_n^{(w)} = W_n^{(1)}/W_n^{(0)}$, then the similarity value for the salient object is defined as:

$$GS_n = \frac{2R_n^{(w)}R_n^{(h)} + C}{(R_n^{(w)})^2 + (R_n^{(h)})^2 + C} \cdot \frac{2S_n^{(0)}S_n^{(1)} + C}{(S_n^{(0)})^2 + (S_n^{(1)})^2 + C} \quad (13)$$

and the global similarity value (GSV) will be the weighted sum according to the absolute size:

$$GSV = \frac{\sum_n GS_n \cdot S_n^{(0)}}{\sum_n S_n^{(0)}} \quad (14)$$

4.3. Information Loss (IL)

An ideal image retargeting method should achieve an optimal balance between the information loss and geometric distortion. In order to measure the information loss, we use the saliency information loss ratio:

$$ILR = 1 - \frac{\sum_k M_k \cdot S_k}{\sum_k S_k} \quad (15)$$

where M_k is the mask map representing the content preservation and S_k is the saliency map. Then, a logistic function is employed to map this ratio to the perceptual information loss (IL):

$$IL = \frac{1}{1 + e^{-\gamma(ILR - \beta)}} \quad (16)$$

	mean KRCC	std KRCC	p-value
BDS	0.083	0.268	0.017
BDW	0.046	0.181	0.869
SIFTflow	0.145	0.262	0.031
EMD	0.251	0.272	1e-5
IR-SSIM	0.363	0.271	1e-3
PGD+SLR	0.415	0.183	9e-14
LGI w/o LSV	0.234	0.301	2.8e-05
LGI w/o GSV	0.349	0.262	4.0e-11
LGI w/o IL	0.382	0.287	6.5e-10
LGI	0.425	0.268	2.6e-09

Table 1: Performance of different metrics on the RetargetMe

4.4. Overall quality

Finally, the retargeted image quality is characterized by the overall similarity score, LGI, which is the product of LSV, GSV and IL:

$$LGI = LSV \cdot GSV \cdot IL \quad (17)$$

4.5. Experimental Results

In order to validate the effectiveness of our metric, we test using the RetargetMe dataset [1]. Using a smaller set of images, the parameters of our metric are determined as: $k_1 = 100$, $k_2 = 1200$, $\gamma = 10$ and $\beta = 0.65$. The proposed metric LGI is compared with BDS [9], BDW [5], SIFTflow [11], EMD[12], IR-SSIM [13] and PGD+SLR [2]. The correlation between the objective and subjective scores for each image is measured by the Kendall Rank Correlation Coefficient:

$$KRCC = \frac{n_c - n_d}{0.5n(n-1)} \quad (18)$$

where n is the length of rankings ($n = 8$), n_c and n_d are the number of concordant and discordant pairs over all pairs, respectively.

Table 1 shows that, our LGI metric ranks the best according to average KRCC. We also analyze the influence of LSV, GSV and IF factors. From Table 1, we can see that these three factors all benefit to the ranking performance. Moreover, the performance of the metric without LSV or GSV deteriorates badly, which correlates the fact that the perception of geometric distortions has more sensitivity than the information loss.

5. CONCLUSION

In this paper, we first propose a halfway domain optimization method for registration with the constraint of SURF pairs, which can robustly find the correspondence under content removal. Then, we propose a LGI metric which measures the local, global similarity, and salient information loss due to image retargeting. A novel edge directional change measure is introduced for local similarity assessment. Experiments show that our metric outperforms the other six methods on the RetargetMe dataset.

References

- [1] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir, “A comparative study of image retargeting,” in *ACM transactions on graphics (TOG)*. ACM, 2010, vol. 29, p. 160.
- [2] Chih-Chung Hsu, Chia-Wen Lin, Yuming Fang, and Weisi Lin, “Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 377–389, 2014.
- [3] Michael Rubinstein, Ariel Shamir, and Shai Avidan, “Improved seam carving for video retargeting,” in *ACM transactions on graphics (TOG)*. ACM, 2008, vol. 27, p. 16.
- [4] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg, “Shift-map image editing,,” in *ICCV*, 2009, vol. 9, pp. 151–158.
- [5] Michael Rubinstein, Ariel Shamir, and Shai Avidan, “Multi-operator media retargeting,” in *ACM Transactions on Graphics (TOG)*. ACM, 2009, vol. 28, p. 23.
- [6] Lior Wolf, Moshe Guttman, and Daniel Cohen-Or, “Non-homogeneous content-driven video-retargeting,” in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–6.
- [7] Philipp Krähenbühl, Manuel Lang, Alexander Hornung, and Markus Gross, “A system for retargeting of streaming video,” in *ACM Transactions on Graphics (TOG)*. ACM, 2009, vol. 28, p. 126.
- [8] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee, “Optimized scale-and-stretch for image resizing,” in *ACM Transactions on Graphics (TOG)*. ACM, 2008, vol. 27, p. 118.
- [9] Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani, “Summarizing visual data using bidirectional similarity,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
- [10] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 416–423.
- [11] Ce Liu, Jenny Yuen, and Antonio Torralba, “Sift flow: Dense correspondence across scenes and its applications,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 978–994, 2011.
- [12] Ofir Pele and Michael Werman, “Fast and robust earth mover’s distances,” in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 460–467.
- [13] Yuming Fang, Kai Zeng, Zhou Wang, Weisi Lin, Zhi-jun Fang, and Chia-Wen Lin, “Objective quality assessment for image retargeting based on structural similarity,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, no. 1, pp. 95–105, 2014.
- [14] Yabin Zhang, Weisi Lin, Xinfeng Zhang, Yuming Fang, and Leida Li, “Aspect ratio similarity (ars) for image retargeting quality assessment,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1080–1084.
- [15] Jing Liao, Rodolfo S Lima, Diego Nehab, Hugues Hoppe, Pedro V Sander, and Jinhui Yu, “Automating image morphing using structural similarity on a halfway domain,” *ACM Transactions on Graphics (TOG)*, vol. 33, no. 5, pp. 168, 2014.
- [16] Mark Finch, John Snyder, and Hugues Hoppe, “Freeform vector graphics with controlled thin-plate splines,” in *ACM Transactions on Graphics (TOG)*. ACM, 2011, vol. 30, p. 166.
- [17] Martin A Fischler and Robert C Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [18] Piotr Dollár and C Lawrence Zitnick, “Structured forests for fast edge detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1841–1848.
- [19] Yuming Fang, Zhenzhong Chen, Weisi Lin, and Chia-Wen Lin, “Saliency-based image retargeting in the compressed domain,” in *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 1049–1052.
- [20] Richard Nock and Frank Nielsen, “Statistical region merging,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 26, no. 11, pp. 1452–1458, 2004.
- [21] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, and Shipeng Li, “Salient object detection: A discriminative regional feature integration approach,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2083–2090.