# A Content-aware Metric for Stitched Panoramic Image Quality Assessment

Anonymous ICCV submission

Paper ID 2428

## Abstract

*One key enabling component of immersive VR visual experience is the construction of panoramic images—each stitched into one large wide-angle image from multiple smaller viewpoint images captured by different cameras. To better evaluate and design stitching algorithms, a lightweight yet accurate quality metric for stitched panoramic images is desirable. In this paper, we design a quality assessment metric specifically for stitched images, where ghosting and structure inconsistency are the most common visual distortions.*

*Specifically, to efficiently capture these distortion types, we fuse a perceptual geometric error metric and a local structure-guided metric into one. For the geometric error, we compute the local variance of optical flow field energy between the distorted and reference images. For the structure-guided metric, we compute intensity and chrominance gradient in highly-structured patches. The two metrics are content-adaptively combined based on the amount of image structures inherent in the 3D scene. Extensive experiments are conducted on our stitched image quality assessment (SIQA) dataset, which contains 408 groups of examples. Results show that the two parts of metrics complement each other, and the fused metric achieves 94.36% precision with the mean subjective opinion. Our SIQA dataset is made publicly available as part of the submission.*

## 1. Introduction

Recent rapid development of *virtual reality* (VR) technologies has led to new immersive visual experiences, rendered using head-mounted displays like Occulus Rift. Real-time reconstruction of panoramic images is one key enabling component, where multiple small viewpoint images captured by an arrangement of cameras on the rig are stitched together into one large wide-angle view [2, 24, 25, 11]. The stitching process can be broadly divided into two parts: i) geometric alignment, and ii) photometric correction. Geometric alignment rectifies the perspectives of the viewpoint images via homographic transformation [13], where the transform parameters (*e.g.*, scaling, rotation, shearing, etc) are computed by establishing correspondence between features in two images' overlapping spatial regions. Thus errors in this stage are primarily caused by the inaccuracy in estimated homographic transform parameters. This results in commonly observed *ghosting* and *structure inconsistency* visual artifacts, as shown in Fig.1.

Photometric correction targets errors due to heterogeneous imaging hardware or environmental conditions among the capturing cameras. Typical errors include vignette and exposure unevenness, which can be removed effectively using a number of post-processing techniques in the literature, including [4, 8, 7]. We thus focus on distortions due to inaccurate estimation of homographic transform parameters in this paper.

Among the diversity of stitching algorithm literatures, many researchers choose to assess the stitched images by either making comparisons subjectively [24, 14] or using conventional image quality assessment (IQA) metrics [12, 1]. However, the problem of stitched image quality assessment (SIQA) differs from classical IQA in two main aspects. First, stitched image quality suffers severely from perspective, scaling and translation distortions, for which conventional IQA methods do not account. Second, instead of the globally diffused noise widely studied in the previous IQA works, the quality of stitched images is more affected by local artifacts such as shape distortion and ghosting introduced by blending surrounding pixels.

**Contributions:** We propose to combine a perceptual geometric error metric and a local structure-guided IQA metric to form a new SIQA metric. To measure the geometric errors, we compute the local variance of optical flow field energy between the distorted and reference images. To measure the structure errors, we compute the intensity and chrominance gradient in highly-structured patches. The two metrics are combined in a content-adaptive manner, where the amount of image structure is first estimated from the originally captured viewpoint images, as illustrated in Fig. 2. Experimental results show that the two parts of metrics complement each other, and the fused metric achieves 94.36% precision with the mean subjective opinion. We

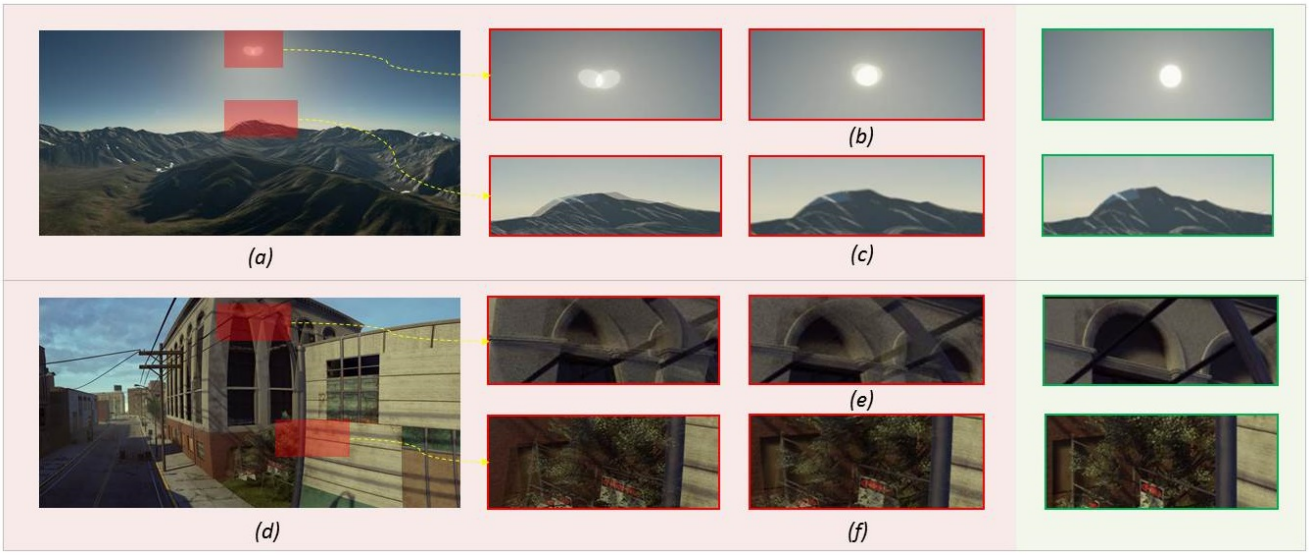


Figure 1. Examples of typical distortions in stitched images. (a) textured scene with ghosting; (b) and (c) are ghosted areas with varying intensity of distortion, distorted image is in red frame, and reference is in green; (d) structured scene with shape breakage; (e) and (f) are the local areas with distorted structure.

also introduce a stitched image quality assessment (SIQA) dataset, which contains 408 groups of examples with perspective variations, which is made publicly available as part of the submission.

The paper is organized as follows. Section 2 discusses previous works in stitched image quality assessment. Section 3 introduces our proposed metric. Experimental results are presented in section 4, and section 5 draws the conclusion.

## 2. Related Work

Compared with the rapid evolution of stitching algorithms in the last decade, previous literatures on SIQA seems insufficient and lagged. The recent applications of the stitching technique has redirected its emphasis, with the auto-adaptive cameras and freely-assembled rigs generalized, the imaging condition has largely been improved and photometric errors introduced on the hardware-level become less a concerning issue. Meanwhile, the demand for VR experience increases the demand for high quality, full-perspective panorama in super resolution.

**Stitching algorithm evaluations.** For stitching algorithms, ghosting and structure inconsistency artifacts that cause large perceived errors and visual discomfort are major challenges [21, 3]. To evaluate how effective the algorithms are as to resolving such errors, many literatures choose to directly compare the stitched images and judge perceptually [24, 14]. The illustration is straight-forward but subjective, and in many cases the comparison is conducted on limited number of examples, which makes the evaluation less convincing. Another way to evaluate stitching algorithm is to adopt classical IQA metrics to stitched images [1, 12] such as MSE (Mean Squared Error) [23], PSNR ((Peak Signal-to-Noise Ratio) [17], SSIM (Structural Similarity index) [6] and VSI (Visual Saliency Induced index) [26]. These are powerful metrics in conventional image quality evaluations, and can effectively grade images generated by global noise addition or various encoding methods, but not designed for the problem of SIQA.

**Previous SIQA metrics.** Much previous SIQA metrics payed more attention to photometric error assessment [10, 13, 22] rather than geometric errors. In [10] and [22], geometric error assessment is omitted and the metrics focus on color correction and intensity consistency. [13] try to quantize the geometric error by computing the structure similarity index (SSIM) of high-frequency information of the stitched and unstitched image difference in the overlapping region. However, since unstitched images used for test are directly cropped from the reference and have no perspective variations, the effectiveness of the method is not proved. In [5] an omni-directional camera system of full-perspective is considered, but the work pays more attention to assessing video consistency among subsequent frames and only adopted a luminance-based metric around the seam. In [16], the gradient of the intensity difference between the stitched and reference image is adopted to assess the geometric error, however, the experiments are conducted on mere 6 stitched image examples, and more experiments are conducted on conventional IQA datasets, which avoids the important and dwells on the trivial.

**IQA-related datasets.** The absence of an SIQA dataset benchmark is another evidence of the problem being understudied. Compared with the popularity of conventional IQA datasets like LIVE database[15] or JPEG 2000[9], the

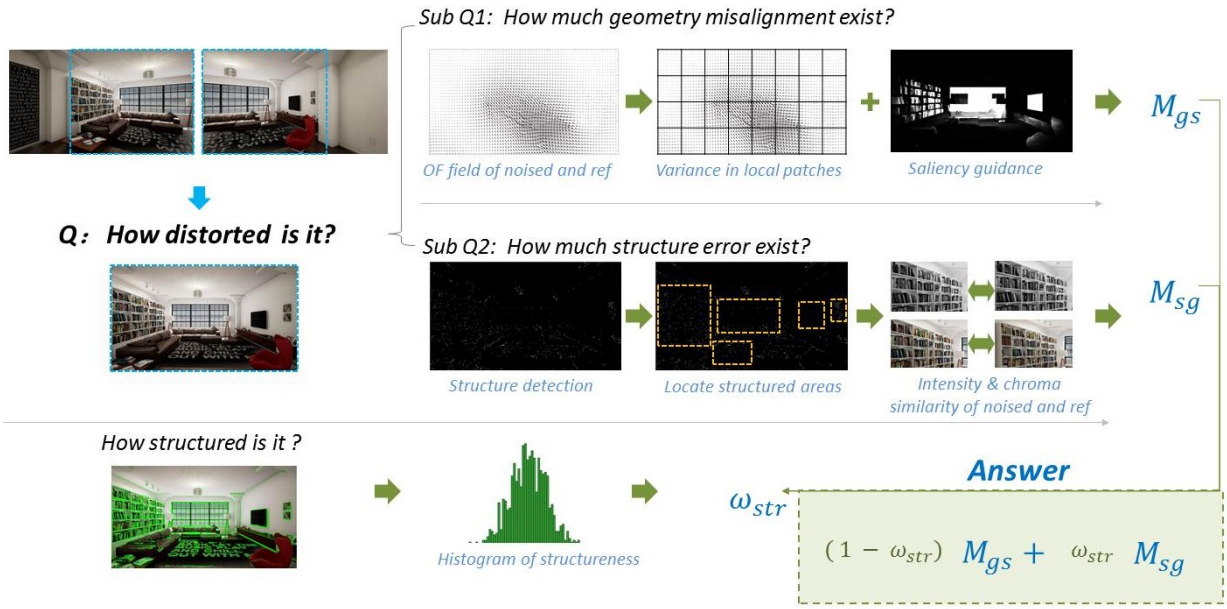


Figure 2. The proposed procedure for stitched image quality assessment.

situation for SIQA problem is obviously a drawback for the development of stitching algorithms. Therefore, establishing a stitched image dataset of proper scale and formation is clearly a necessary move.

## 3. Proposed Method

**Perceptual geometric error metric.** As mentioned earlier, miscalculated correspondence between the unstitched viewpoint images is a major source of distortion for stitched images that results in relative perspective, scale and translation error. To estimate such errors, a perceptual geometric error metric is proposed. First, we establish a dense correspondence between the stitched and reference images to identify the transformation at the pixel level using optical flow. Considering the diversity of existing stitching algorithms, displacement between the stitched and reference images might vary across spatial dimensions. Thus large displacement optical flow (LDOF) [20] is adopted to calculate point correspondence. The dense flow field is then obtained as motion vectors at each pixel, which is later utilized to assemble the geometric error metric.

The magnitude of flow field reflects the intensity of geometric transformation from the stitched image to the reference image. However, what characterizes geometric distortions is the relative perspective, scale and translation variations, which are found in the local patches. Hence, the variance of flow in an $N$-by-$N$ local patch is adopted to describe local geometric error. The error metric $M_g$ for each stitched image is then obtained by summing up the variance of local patches as follows:

$$M_g = \sum_P \left( \frac{1}{N^2 - 1} \sum_{i=1}^{N^2} |g_i - \mu_i|^2 \right) \quad (1)$$

where $P$ is the number of patches, $N$ is the patch size and $\mu_i$ is the mean magnitude within the $i^{th}$ patch.

Although the distribution of geometric errors is characterized as random, how human perceives the error is quite structure-based. For stitched image with broad filed of view providing rich information, human visual perception has more impact on how 'terrible' the distortion means to its viewer. To this end, a salient object detection model is applied to generate an attention-weighted map for each reference image. Thus, the saliency guided geometric error metric $M_{gs}$ is summarized as follows:

$$M_{gs} = \sum_P S_p \cdot \left( \frac{1}{N^2 - 1} \sum_{i=1}^{N^2} |g_i - \mu_i|^2 \right) \quad (2)$$

where $S_p$ is the normalized saliency within the $p^{th}$ patch.

**Structure-guided metric.** Despite the measurement of miscalculated correspondence using geometric error metric, shape and chrominance similarity are proven effective means to assess noticeable structure distortions [16, 26]. To adopt the measurement properly, we use stitched images with rectified perspectives. Furthermore, given the difference between SIQA distortions and conventional IQA distortions, the former is more structure-based and the latter is more globally diffused. Hence we customize a structure-guided metric for SIQA problems. First, the structured areas are located as bounding-boxes, then the visual saliency

index (VSI) [26] is applied to each bounding-box. VSI is an effective metric combining visual saliency, edge similarity and chrominance consistency, which is in accordance with the desired measurement. Finally, we sum the index along the bounding-boxes to form the metric.

We rectify the geometric differences by warping the stitched image to the reference image using the calculated LDOF field. The structured areas are located using the line segment detector (LSD) [19] method, and a bounding-box is imposed around each line with sufficient length. Thus the structure-guided metric $M_{sg}$ is presented as follows:

$$M_{sg} = \sum_{B} S_b \tag{3}$$

where $S_b$ is the VSI score for the $b^{th}$ bounding box in the structured area.

The geometric error metric quantifies the misalignment, and hence is suitable for texture-oriented distortions like ghosting. On the other hand, the structure-guided metric characterizes the shape and color inconsistency. As a result, it is necessary to first decide how structured a scene is before the two components are combined. To this end, we design a metric that quantifies the "structureness" of a scene. Lines are segmented using the LSD method and pooled into a 30-dimension histogram according to their phase, the magnitude for each bin is computed as follows:

$$B_{mag} = \left(\frac{\mu}{\sigma + \epsilon}\right) \sum_{Q} \exp^{L_q/\gamma} \tag{4}$$

where $\mu$ and $\sigma$ are the respective mean and variance of line length in that bin, $Q$ is the number of lines, and $L_q$ is the length of $q^th$ line within the bin. $\gamma$ is the rectification parameter; in this paper we use one-tenth of the diagonal length for each stitched image. The structureness index is described as follows:

$$\omega_{str} = \sum_{i=1}^{B} B_{mag} + \sum_{i=1}^{B_{top}} B_{mag} \tag{5}$$

where $B$ is the number of bins, 30 bins are divided in our experiment and $B_{top}$ is the number of bins with top magnitude and in this paper we adopt 5 as $B_{top}$. The structureness index is normalized between $[0, 1]$ using the min-max method and then further rectified with the normal cumulative distribution. Fig. 3 illustrates typical examples computing structureness.

Finally, the content-aware adaptive metric is composed as follows:

$$M = \omega_{str} \cdot M_{sg} + (1 - \omega_{str}) \cdot M_{gs} \tag{6}$$

.

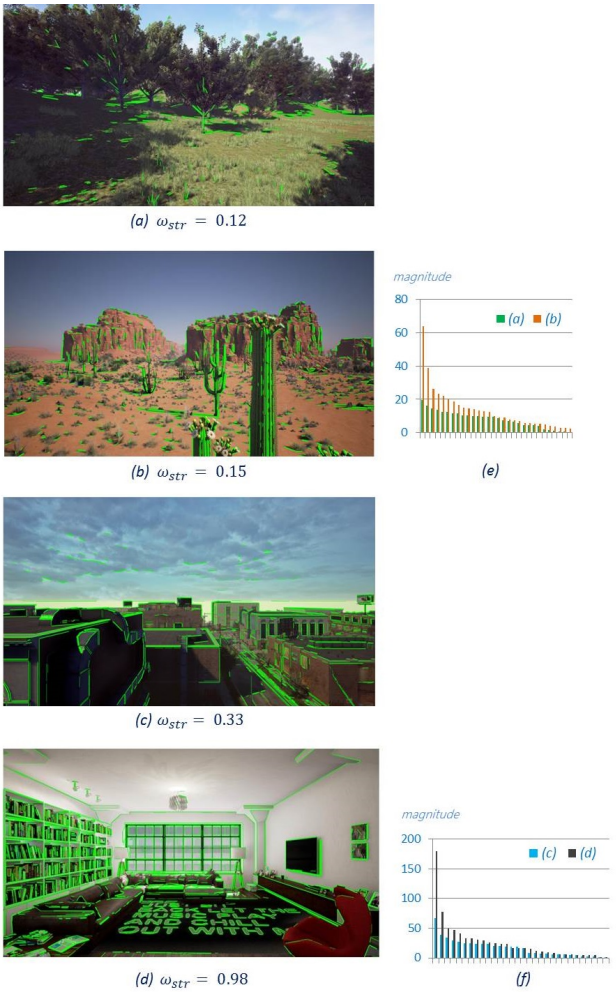


Figure 3. Examples of computing structureness index $\omega_{str}$. (a) is a natural textured scene with relatively less structure; (b) is a natural scene with more structure; (c) is an outdoor structured scene; (d) is a structured indoor scene with high structureness index.

## 4. Experimentation

In this paper, we introduce a stitched image quality assessment dataset benchmark called SIQA dataset. Extensive experiments are conducted on the SIQA dataset, including the comparison between our proposed metric and classical IQA metrics, the validation of each metric component, and the contrast between fixed-weight and content-aware adaptive combination mechanism. To analyze the combined metric and how each component takes effect, we also studied the specific examples using each component solely. Results show the effectiveness of the proposed content-aware metric, achieving 94.36% precision compared with the mean subjective opinion score (MOS).

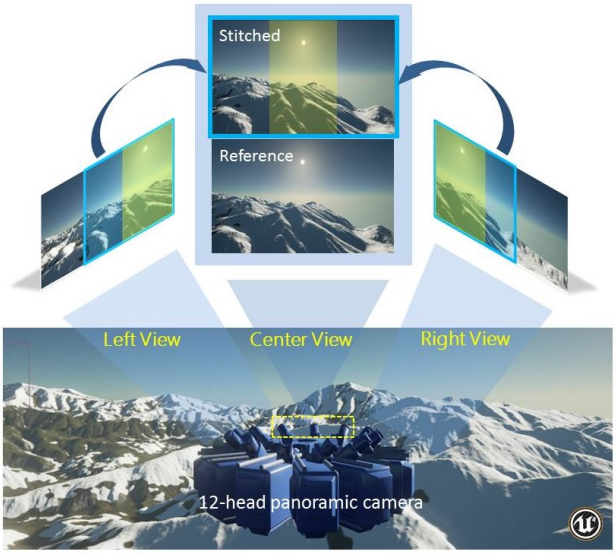


Figure 4. The 12-head panoramic camera established in a virtual scene using the Unreal Engine, and the formation of stitched/reference image pairs for SIQA-dataset.

### 4.1. SIQA Dataset Benchmark

The first version of our SIQA dataset is based on synthetic virtual scenes, since we try to evaluate the proposed metric for various stitching algorithms under ideal photometric conditions. The images are obtained by establishing virtual scenes with the powerful 3D model tool—Unreal Engine. A synthesized 12-head panoramic camera is placed at multiple locations of each scene, covering 360 degree surrounding view, and each camera has an FOV (field of view) of 90 degree. Exactly one image is taken for each of the 12 cameras at one location simultaneously. Each camera view is used as a full reference of the stitched view of its left and right adjacent cameras, as demonstrated in Fig.4.

SIQA dataset utilized twelve different 3D scenes varying from wild landscapes to structured scenes, two sets of stitched images are obtained using a popular off-the-shelf stitching tool Nuke using different parameter settings, altogether 816 stitched samples, the original images are in high-definition with $3k - by - 2k$ in size. Annotations from 28 different viewers are integrated to decide on which one of the two stitched images is better, more than 10000 decisions are combined into mean subjective opinion (MOS), which we later utilize as the ground-truth.

The dataset is properly constructed both in formation and in scale, and to the best of our knowledge, is also the first stitched image dataset considering perspective variations.

| Metric | Precision with MOS |
|---|---|
| VSI | **0.8701** |
| SSIM | 0.8162 |
| FSIM | 0.8162 |
| GSM | 0.8407 |
| SR-SIM | 0.8333 |
| RF-SIM | 0.6691 |
| **Proposed** | **0.9436** |

Table 1. Comparison of the proposed metric with the classical IQA metrics, best results for classical IQA metrics and for all the evaluated metrics are high-lighted in bold text.

| Metric | Without | Quarter-size | Half-size | Origin-size |
|---|---|---|---|---|
| $M_{gs}$ | 0.7034 | 0.7696 | 0.7770 | **0.7868** |

Table 2. The saliency fineness and the correspondingly achieved precision.

The dataset is formed in a structured way, each group of captures by the virtual camera rig is separately provided that researchers are enabled to choose different size of overlap for stitching algorithm evaluations.

### 4.2. Experimental Results

We mainly conducted 3 groups of experiments on the SIQA dataset, including comparing our proposed metric with the classical IQA metrics, evaluating the effectiveness of each metric component and validating the effectiveness of the combination mechanism.

Six conventional IQA metrics are evaluated solely comparing with the proposed metric, as illustrated in Tab.1. Evaluated as single metric, VSI has better performance compared to others, yet the overall precision is unsatisfying since they are not designed for the stitched image evaluation problem.

For saliency detection, we adopted a Minimum-Spanning-Tree-based (MST) [18] method which is both effective and basically real-time. As mentioned in the previous section, the calculated saliency magnitude is summed-up and normalized in local patches and then used as the perceptual weight. As much previous work suggested, saliency guidance serves a positive but non-dominant role in IQA-related problems. Meanwhile, it is observed that the fineness of the saliency map is positively-correlated, as illustrated in Tab.2.

The structure-guided metric is obtained by computing intensity and chrominance gradient around local structured patches. As mentioned earlier, the structured areas are located by lines detection using the LSD method. To eliminate the trivial areas and focus on the real structureness, we remove the detection results with length below average. Finally, VSI is computed in each bounding box imposed

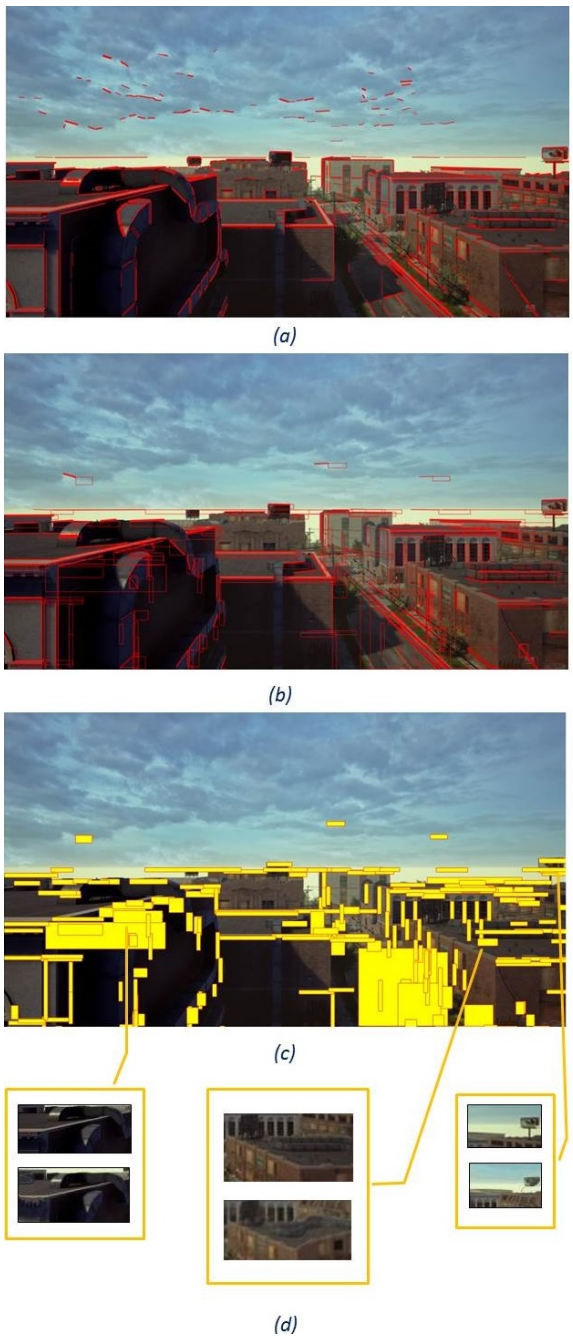


Figure 5. An example of structure guidance for computing local image quality assessment. (a) is image after LSD detection, the red lines are the detection results; (b) is the result after trivial structures being removed; (c) is the image with bounding boxes of structured area, high-lighted in yellow; (d) is the amplified examples of bounding box.

around each reserved line. Fig.5 illustrates a typical example under this process.

In the last section, we propose to adaptively combine the geometric error metric and structure-guided metric according to scene structureness. To validate the proposed idea, contrast experiments are conducted including using the geometric error metric and structure-guided metric solely, combining them with a fixed-weight mechanism, and using the content-adaptive way. The fixed weight adopted in this experiment is 0.5 and 0.5. As illustrated in Tab.3, the results show that combining the two components promotes the precision of the assessment, and best result is achieved using the content-aware adaptive combination, hitting 94.36% precision with the MOS.

| Metric component | Precision with MOS |
|---|---|
| $M_{gs}$ | 0.7868 |
| $M_{sg}$ | 0.9167 |
| Fixed combine | 0.9216 |
| Content-aware combine | **0.9436** |

Table 3. The saliency fineness and achieved precision.

Though comparisons clearly reveal the effectiveness of the proposed method, we still need to validate that the two components are practically complementary to each other. To this end, a close observation is conducted among the examples for which one component works but another one fails. Part of the examples are illustrated in Fig.6, we observe that in unstructured scenes like (a) when two stitched images have very similar structure, even similar distortions, attention-based IQA metric fails while geometric error metric successfully scored image 1 higher since the geometric distance error between image 1 and reference is relatively smaller. In structured scenes like (b) where diverse edge breakage and shape distortion exist, geometric error metric fails to evaluate the differences while the structure-guided metric successfully captured the distorted areas, thus providing better decisions. Based on observation through such examples, the correctness of our previous conception that the two component complement each other shows.

## 5. Conclusion

We propose a quality assessment metric specifically designed for stitched images. We first analyze different error types typically encountered in image stitching, including how the errors are generated and rendered, and then arrive at the most common visual distortions in SIQA—ghosting and structure inconsistency. To effectively characterize these distortion types, we propose to adaptively fuse a perceptive geometric error metric and a structure-guided metric.

To capture perceptual ghosting which is mostly caused by geometric misalignment, we compute the local variance of optical flow field energy between the distorted and reference images, guided by detected saliency. For structure inconsistency, a powerful intensity and chrominance gradient

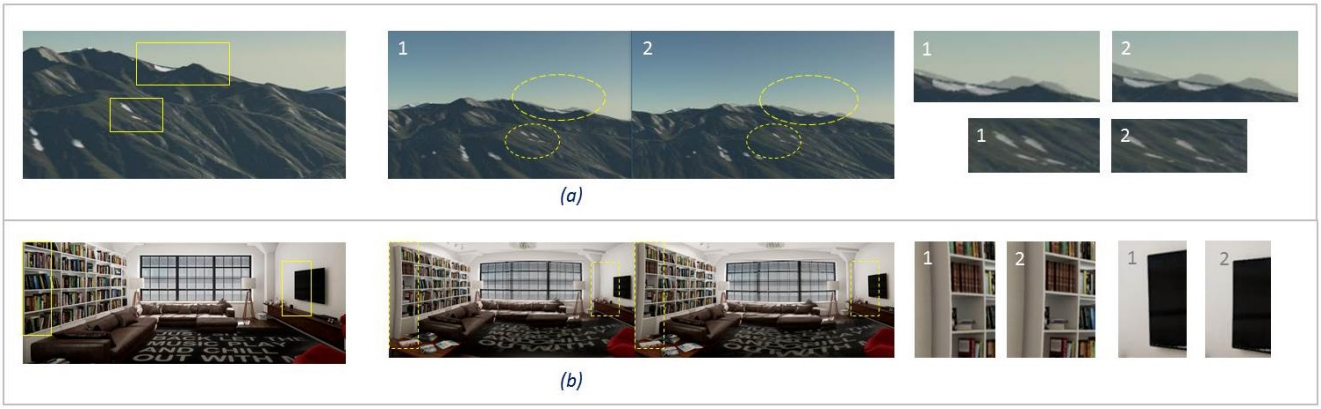


Figure 6. Examples of the two metric components complement each other. (a) is the example that geometric error metric score the stitched image 1 higher, yet the local structure-guided metric score image 1 lower; (b) is the example which structure-guided metric score image 1 higher but the geometric error metric vice versa.

index VSI is adopted and customized around the highly-structured areas of the stitched images. Based on understanding of the different purposes of these two metrics, we propose to use a content-adaptive combination according to the specific scene structure. Experimental results show the effectiveness of our proposed metric and confirm the correctness of the combination mechanism. The metric can be used to optimize various stitching algorithms.

Extensive experiments are conducted using our SIQA dataset, which we introduce as a dataset benchmark for SIQA problems. The large-scale dataset is laboriously constructed and is made publicly available for researchers in the VR community for further research.

## References

[1] E. Adel, M. Elmogy, and H. Elbakry. Image stitching based on feature extraction techniques: a survey. *International Journal of Computer Applications (0975-8887) Volume*, 2014. 1, 2

[2] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007. 1

[3] C.-H. Chang, Y. Sato, and Y.-Y. Chuang. Shape-preserving half-projective warps for image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3254–3261, 2014. 2

[4] M. Harville, B. Culbertson, I. Sobel, D. Gelb, A. Fitzhugh, and D. Tanguay. Practical methods for geometric and photometric correction of tiled projector. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 5–5. IEEE, 2006. 1

[5] S. Leorin, L. Lucchese, and R. G. Cutler. Quality assessment of panorama video for videoconferencing applications. In *Multimedia Signal Processing, 2005 IEEE 7th Workshop on*, pages 1–4. IEEE, 2005. 2

[6] L. Liu, H. Dong, H. Huang, and A. C. Bovik. No-reference image quality assessment in curvelet domain. *Signal Processing: Image Communication*, 29(4):494–505, 2014. 2

[7] Y. Liu and B. Zhang. Photometric alignment for surround view camera system. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1827–1831. IEEE, 2014. 1

[8] S. Lu and C. L. Tan. Thresholding of badly illuminated document images through photometric correction. In *Proceedings of the 2007 ACM symposium on Document engineering*, pages 3–8. ACM, 2007. 1

[9] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011. 2

[10] P. Paalanen, J.-K. Kämäräinen, and H. Kälviäinen. Image based quantitative mosaic evaluation with artificial video. In *Scandinavian Conference on Image Analysis*, pages 470–479. Springer, 2009. 2

[11] F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross. Panoramic video from unstructured camera arrays. In *Computer Graphics Forum*, volume 34, pages 57–68. Wiley Online Library, 2015. 1

[12] Y. Qian, D. Liao, and J. Zhou. Manifold alignment based color transfer for multiview image stitching. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 1341–1345. IEEE, 2013. 1, 2

[13] H. Qureshi, M. Khan, R. Hafiz, Y. Cho, and J. Cha. Quantitative quality assessment of stitched panoramic images. *IET Image Processing*, 6(9):1348–1358, 2012. 1, 2

[14] C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung. Megastereo: Constructing high-resolution stereo panoramas. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1256–1263, 2013. 1, 2

[15] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2. 2005. 2

[16] M. Solh and G. AlRegib. Miqm: A novel multi-view images quality measure. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pages 186–191. IEEE, 2009. 2, 3

[17] A. Tanchenko. Visual-psnr measure of image quality. *Journal of Visual Communication and Image Representation*, 25(5):874–878, 2014. 2

[18] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien. Real-time salient object detection with a minimum spanning tree. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2334–2342, 2016. 5

[19] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32(4):722–732, 2010. 4

[20] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1385–1392, 2013. 3

[21] T. Xiang, G.-S. Xia, and L. Zhang. Image stitching with perspective-preserving warping. *arXiv preprint arXiv:1605.05019*, 2016. 2

[22] W. Xu and J. Mulligan. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 263–270. IEEE, 2010. 2

[23] W. Xue, L. Zhang, X. Mou, and A. C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 2

[24] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter. As-projective-as-possible image stitching with moving dlt. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2339–2346, 2013. 1, 2

[25] F. Zhang and F. Liu. Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269, 2014. 1

[26] L. Zhang, Y. Shen, and H. Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. 2, 3, 4