GC360IQ: Gradient-Detail Consistency Model for 360-degree Stitched Image Quality Assessment

Jinghan Zhou
University of Waterloo
j263zhou@uwaterloo.ca

Zhou Wang
University of Waterloo
zhou.wang@uwaterloo.ca

Abstract

Evaluating the perceptual quality of 360° stitched panoramas is challenging because no pristine reference exists and the distortions are localized along stitching seams rather than global. Conventional full-reference and no-reference IQA methods are mainly designed for compression or projection artifacts and thus fail to address stitching-specific degradations such as luminance inconsistency and detail loss. We propose GC360IQ, a perception-referenced framework that assesses stitching quality based on the human visual expectation of luminance continuity and structural fidelity. A dedicated stitched-image database is constructed to isolate blending-induced luminance inconsistency and detail loss while minimizing geometric misalignment. These two perceptual dimensions are modeled using dual CNN branches inspired by gradient and structural features, whose fused representation predicts overall perceptual quality. Subjective experiments using an HMD with 30 participants show that GC360IQ achieves significantly higher correlation with mean opinion scores than existing FR and NR IQA methods. The framework offers an interpretable, stitching-aware solution and provides guidance for perceptually optimized blending.

1. Introduction

In virtual reality (VR), 360° stitched panoramas are widely used to create immersive visual experiences. They are generated by stitching multiple camera views into a single spherical image, where alignment and blending determine the overall perceptual consistency. The perceptual quality of such panoramas directly affects visual experience, while reliable assessment remains difficult because no pristine reference image exists. The stitching process, which involves geometric alignment, color correction, and multi-view blending, inevitably introduces spatially localized distortions that differ from the global degradations considered in conventional full-reference (FR) and no-reference (NR) image quality assessment (IQA) methods.

Recent perceptual psychology research shows that immersive visual experience depends not only on sensory fidelity but also on observers' expectations of realism and motion coherence [10]. This supports the perceptual-referenced view that human judgments of image quality are guided by expectations of luminance and structural consistency rather than pixel-level fidelity. Previous studies [7, 21] have shown that among various stitching-induced artifacts, *luminance inconsistency* and *detail loss* are two types of distortions that have a strong influence on subjective quality. Luminance inconsistency often occurs when exposure differences between adjacent views are not smoothly compensated across the seam, while detail loss results from over-smoothing during blending. To address these perceptual issues, we propose a model that directly reflects their influence on perceived image quality rather than treating them as general degradations.

In this work, we present **GC360IQ**, a perception-referenced framework for assessing the quality of stitched 360° panoramas. Our main contributions are as follows: **1)** We introduce a perceptual reference paradigm that evaluates stitching quality based on perceived visual consistency rather than pixel fidelity, aligning the task with human perceptual mechanisms. **2)** We build the first stitched 360° image database that isolates blending distortions: luminance inconsistency and detail loss, while minimizing geometric misalignment. **3)** We design an interpretable feature fusion model that combines gradient-based [5] and structure-based [17] features to predict perceptual quality. **4)** We conduct extensive subjective tests and comparisons with state-of-the-art FR and NR methods, showing that GC360IQ achieves the highest correlation with human opinion and supports perceptually optimized blending in future work.

2. Literature Review

Traditional IQA methods fall into two classes: full-reference (FR) and no-reference (NR). FR metrics such as SSIM [17] and GMSD [20] measure structural or gra-

dient similarity to a reference, while NR metrics such as BRISQUE [8] and NIQE [9] estimate quality from natural-scene statistics or learned perceptual features.

For 360° images, FR models like WS-PSNR [18] and WS-SSIM [23] adapt pixel-level measures to the equirectangular projection. Recent NR models like VGCN [19] and ST360IQ [16] use deep networks to predict perceptual quality directly from distorted panoramas. These methods correlate well with human opinion but depend on large labeled datasets and saliency cues, and mainly address global distortions such as compression or projection rather than stitching-induced local artifacts.

Existing panoramic IQA datasets, including CVIQD [12] and OIQA [3], focus on global degradations. In contrast, Yang *et al.* [21] and Madhusudana et al. [7] introduced stitched-image datasets capturing four artifact types—luminance inconsistency, texture detail loss, geometric misalignment, and structural discontinuity. The first two dominate perceptual judgments and stem from blending operations. However, these studies lack perception-based modeling and dimension-wise annotations. GC360IQ fills this gap with an interpretable, perception-referenced framework for stitched 360° image quality assessment.

3. Proposed Method

GC360IQ evaluates stitched 360° image quality through three perceptual dimensions: luminance inconsistency, detail loss, and overall quality.

3.1. Preprocessing

For each scenario, unblended input images and the stitched panorama are processed. Each panorama is divided into seven directional subregions following the equirectangular FOV extraction strategy [6]. Perspective views are cropped from these directions and resized to 256×256 pixels. A corresponding seam-boundary mask is generated using the same projection to define valid seam regions.

Each sample yields $(\mathbf{x}, \mathbf{y}, \mathbf{p})$, where $\mathbf{x}$ and $\mathbf{y}$ are the stitched and unblended patches, and $\mathbf{p}$ is the binary mask restricting feature computation to valid seam areas, eliminating bias from non-overlapping regions.

3.2. Luminance Inconsistency Prediction

Luminance inconsistency between the stitched and unblended patches is estimated from spatial gradients. The patches are convolved with the 11×11 horizontal and vertical derivative filters $\mathbf{h} = \mathbf{u_0}\mathbf{u_1}^\top$ and $\mathbf{v} = \mathbf{u_1}\mathbf{u_0}^\top$ following the Farid–Simoncelli design [5]. The resulting gradient matrices are $\mathbf{S} = [\mathbf{x} * \mathbf{h},\ \mathbf{x} * \mathbf{v}]$ and $\mathbf{O} = [\mathbf{y} * \mathbf{h},\ \mathbf{y} * \mathbf{v}]$, where $*$ denotes 2D convolution with symmetric padding to preserve boundary continuity [17]. A binary mask $\mathbf{p}$ filters seam-relevant pixels.

Luminance inconsistency is characterized by the gradient magnitude and directional differences:

$$a_{m,n} = \|\mathbf{s}_{m,n} - \mathbf{o}_{m,n}\|_2, \quad b_{m,n} = 1 - \frac{\mathbf{s}_{m,n} \cdot \mathbf{o}_{m,n}}{\|\mathbf{s}_{m,n}\|_2 \|\mathbf{o}_{m,n}\|_2}, \tag{1}$$

where $\mathbf{s}_{m,n}$ and $\mathbf{o}_{m,n}$ are local gradient vectors at pixel (m, n). Seven pairs of $\mathbf{A} = \{a_{m,n}\}$ and $\mathbf{B} = \{b_{m,n}\}$ maps from different FOVs are concatenated as $[\mathbf{A}_1, \ldots, \mathbf{A}_7, \mathbf{B}_1, \ldots, \mathbf{B}_7]$ to form a 14-channel tensor $\boldsymbol{\eta} \in \mathbb{R}^{14\times256\times256}$, which is input to a CNN f_{CNN} for luminance inconsistency prediction. The CNN comprises five convolutional and two fully connected layers, with feature depth expanding from 14 to 64 channels via alternating 5×5 and 3×3 kernels (strides $\{2, 1, 2, 1, 2\}$), each followed by ReLU activation and 2×2 max pooling, except for the final linear output layer.

3.3. Detail Loss Prediction

Detail loss is reflected by luminance, contrast, and structural variations between stitched and unblended images. Following SSIM [17], local statistics are computed within an 11×11 window with symmetric padding:

$$l_{m,n} = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c_{m,n} = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{2}$$

where μ and σ denote the local means and standard deviations of the stitched and unblended patches. The corresponding SSIM quality map $d_{m,n}$ is also obtained to capture structural fidelity. A binary mask $\mathbf{P}$ is applied afterward to retain only seam-relevant pixels. For each of the seven FOVs, the luminance, contrast, and SSIM maps are concatenated to form a 21-channel tensor $\boldsymbol{\zeta} \in \mathbb{R}^{21\times256\times256}$ in the same manner as in Sec. 3.2, which is fed into g_{CNN}. The architecture mirrors f_{CNN}, differing only in input channels.

3.4. 360-degree Image Quality Assessment

The final quality score is obtained by fusing intermediate features from the two branches above. Feature vectors from the first fully connected layers, $\mathbf{f}_{\text{fc1}}$ and $\mathbf{g}_{\text{fc1}}$, are concatenated as $[\mathbf{f}_{\text{fc1}}, \mathbf{g}_{\text{fc1}}]$ to form a 128-dimensional feature $\boldsymbol{\Phi} \in \mathbb{R}^{128}$. A three-layer fully connected network t_{FCN} with hidden dimensions 128 and 256 and ReLU activations produces the final perceptual quality score: $Q = t_{\text{FCN}}(\boldsymbol{\Phi})$.

4. Experiments

We evaluate the proposed 360° image quality assessment framework through image capture, subjective testing, CNN training, and performance analysis.

4.1. Image Capture

A seven-camera GoPro Hero4 Black rig (Fig. 1) is used to capture multi-view fisheye images for panoramic stitching.

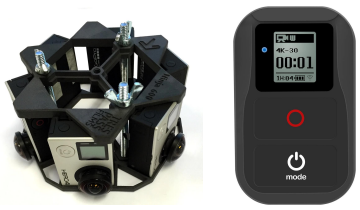


Figure 1. Seven-camera rig setup with GoPro Hero4 Black units. Six cameras capture equatorial views and one captures the zenith.

Six cameras are evenly spaced along the equatorial plane ($\theta = 90°$, $\phi \in \{0°, 60°, \ldots, 300°\}$), and one is oriented upward ($\theta = 0°$) to cover the zenith. Automatic exposure is enabled for each camera so that every view uses its direction-specific optimal exposure level. All cameras are triggered simultaneously via a wireless remote, ensuring precise temporal synchronization and sufficient overlap for stitching.

Twenty panoramic scenarios (10 outdoor daytime, 4 nighttime, and 6 indoor) are captured under diverse illumination conditions. Each scenario is processed using five blending algorithms: Naive Blending, Feather Blending (FB) [14], Multi-Band Blending (MBB) [24], Mean-Value Coordinates Blending (MVCB) [4], and Modified Poisson Blending (MPB) [15], to generate the test images.

To isolate blending-induced distortions, state-of-the-art feature detection, matching, and alignment methods are employed to minimize geometric errors, effectively removing spatial misalignments and allowing the analysis to focus on the perceptual impact of blending on luminance inconsistency and detail loss across seams.

4.2. Subjective Test

A controlled subjective test is conducted using a head-mounted display (HMD), allowing observers to freely explore each panorama in a full 360° environment while seated on a swivel chair. Thirty subjects (17 males and 13 females, aged 18–35) evaluate 100 stitched panoramas (20 scenarios × 5 blending algorithms). The order of scenarios and blending results is randomized to avoid ordering and fatigue bias.

Each image is rated on three perceptual dimensions—*luminance inconsistency*, *detail loss*, and *overall quality*—using a five-point discrete scale (1=*Bad*, 5=*Excellent*) following the Single-Stimulus ACR protocol [1]. Before scoring, subjects view two contextual visualizations inside the HMD: (1) a combined 2D FOV layout showing seven directional views extracted from the unblended inputs, and (2) a 360° seam visualization highlighting visible boundaries. Afterward, the five stitched results of each scenario are presented sequentially for rating relative to the internal visual expectation formed from the unblended views. Higher scores indicate smoother brightness across seams, better texture detail preservation, and higher overall perceptual quality. A short training session precedes the formal test. The obtained three scores per image are used as ground-truth labels for model training (Section 4.3).

All subjective scores s_j are Z-score normalized per participant as $z_j = (s_j - \mu_j)/\sigma_j$ to remove rating bias, where μ_j and σ_j are the mean and standard deviation of scores from subject j. Normalized scores are then rescaled to the original five-point range and outliers are removed following ITU-R BT.500. The final mean opinion score (MOS) is obtained by averaging across subjects: $MOS = \frac{1}{N}\sum_{j=1}^{N} z_j$, where N is the number of valid participants.

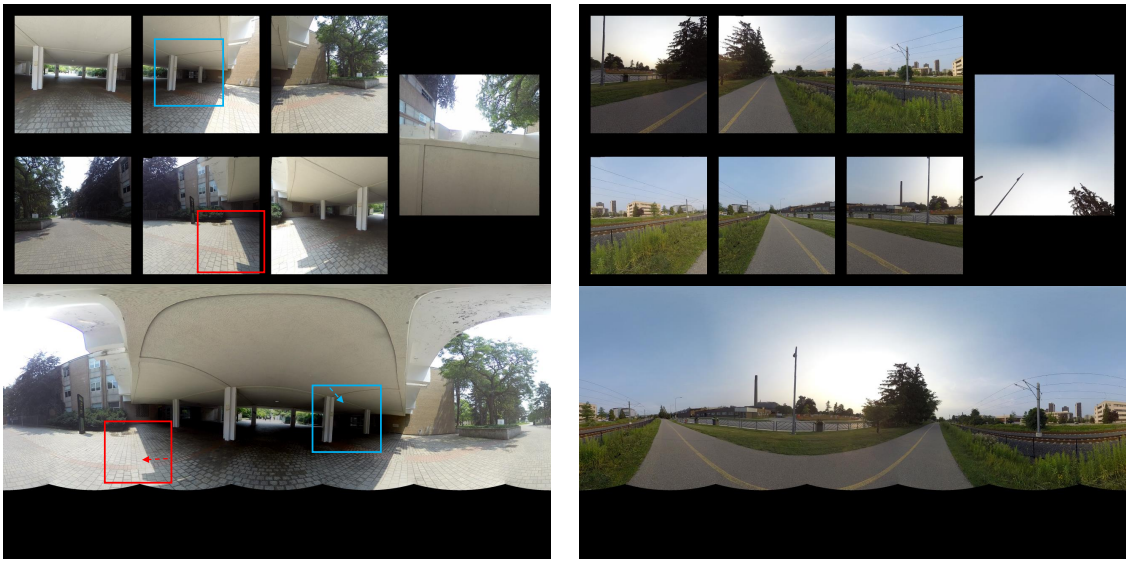
4.3. CNN Training Details

All CNN models are trained and tested using a scenario-wise split of the 20 panoramic scenes, with 10 scenarios for training and 10 for testing across different environment types. Each network predicts its corresponding subjective score (*luminance inconsistency*, *detail loss*, or *overall quality*) obtained from the subjective test. f_{CNN} and g_{CNN} are trained independently with their respective MOS labels, and t_{FCN} is trained on the fused features extracted from the first fully connected layers of the two trained branches. All input tensors are normalized before training, and network weights are initialized using standard PyTorch settings. The Adam optimizer with an initial learning rate of 1×10^{-4} is employed, and training is stopped early if the validation loss does not improve for 10 consecutive epochs. Performance metrics and comparative results are reported in Section 4.4.

4.4. Experimental Results

The proposed method is evaluated on the constructed 360° stitched image database and compared with the state-of-the-art IQA algorithms. Three standard criteria are used: the Pearson Linear Correlation Coefficient (PLCC), the Spearman Rank-Order Correlation Coefficient (SRCC), and the Root Mean Square Error (RMSE). Following the common practice [3], a five-parameter logistic function maps the predicted scores to MOS values before PLCC and RMSE computation using a 50/50 database split for parameter fitting and correlation evaluation to ensure fair comparison. For non-learning models, the random split is repeated 50 times, and the average result is reported. Learning-based models, including GC360IQ, VGCN, and ST360IQ, follow the scenario-wise split in Section 4.3, and each is trained and tested five times with averaged performance reported. When model outputs are inversely correlated with MOS, the absolute SRCC value is used for consistency.

Table 1 presents the quantitative results on our database. **GC360IQ** attains the highest PLCC and SRCC and the lowest RMSE, showing the best alignment with human perceptual judgments of blending-induced luminance inconsis-



(a) (b)

Figure 2. Examples of stitched panoramas with predicted quality scores. (a) DISTS: 3.53, ST360IQ: 3.78, GC360IQ (ours): 2.79. (b) DISTS: 3.45, ST360IQ: 3.06, GC360IQ (ours): 4.17. Red box marks luminance inconsistency and the blue box highlights detail loss. These artifacts are fairly visible in the above figures but become disturbing when viewed in a VR environment.

Table 1. Quantitative comparison of GC360IQ with existing IQA methods on our database. The best results are highlighted in bold.

Type	Methods	PLCC	SRCC	RMSE
NR	NIQE [9]	0.0853	0.1065	0.8851
	BRISQUE [8]	0.0364	0.0406	0.8940
	VGCN [19]	0.4208	0.4282	0.8721
	ST360IQ [16]	0.6192	0.5889	0.7174
FR	PSNR	0.0325	0.1167	0.8870
	SSIM [17]	0.3375	0.0962	0.8161
	WS-PSNR [13]	0.0307	0.1023	0.8838
	WS-SSIM [23]	0.3208	0.0818	0.8310
	FSIM [22]	0.2807	0.2207	0.8277
	VIF [11]	0.6628	0.6734	0.6633
	GMSD [20]	0.3238	0.3544	0.8433
	DISTS [2]	0.6689	0.6801	0.6494
–	**GC360IQ (Ours)**	**0.8555**	**0.8319**	**0.4378**

tency and detail loss. As shown in Fig. 2, image (a) exhibits slight luminance inconsistency in the red box and noticeable detail loss in the blue box, but DISTS and ST360IQ fail to assign low scores. Image (b) is well stitched with smooth luminance and clear details, but the competing methods still underestimated its quality. GC360IQ produces more perceptually consistent predictions in both cases, confirming its effectiveness in modeling stitching-related distortions.

We assess the contribution of each component through three ablation experiments, including removing the feature, the structure branch, and the gradient branch. The results in Table 2 show consistent performance degradation when any component is excluded. This confirms that each perceptual feature and the fusion stage are integral to the model's overall effectiveness.

Table 2. Ablation study evaluating the effect of input features and branch fusion on model performance.

Method	PLCC	SRCC	RMSE
w/o Feature (Image $\times$ Mask only)	0.3586	0.4192	0.9264
w/o Structure Branch (Gradient only)	0.7727	0.7418	0.5394
w/o Gradient Branch (Structure only)	0.8044	0.8027	0.4940
Proposed Model (GC360IQ)	**0.8555**	**0.8319**	**0.4378**

5. Conclusion

In this paper, we presented **GC360IQ**, a perception-referenced framework for assessing the quality of 360° stitched panoramas following a perceptual reference paradigm, where image quality is evaluated relative to the observer's expected visual consistency rather than a pristine reference. By decomposing perceptual quality into two interpretable dimensions: luminance consistency and detail fidelity, the model captures key perceptual cues characterizing blending-related degradations in stitched images. Built upon the first stitched-image database focusing solely on blending distortions and validated through subjective experiments, GC360IQ shows strong agreement with human opinion and outperforms existing FR and NR IQA methods. The results show that modeling luminance inconsistency and detail loss effectively describes the perceptual impact of blending, offering a practical direction for developing perceptually optimized blending algorithms as future work.

References

[1] Methodology for the subjective assessment of the quality of television pictures. ITU-R Recommendation BT.500-13, International Telecommunication Union, Geneva, Switzerland, 2012. 3

[2] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020. 4

[3] Huiyu Duan, Guangtao Zhai, Xiongkuo Min, Yucheng Zhu, Yi Fang, and Xiaokang Yang. Perceptual quality assessment of omnidirectional images. In *2018 IEEE international symposium on circuits and systems (ISCAS)*, pages 1–5. IEEE, 2018. 2, 3

[4] Zeev Farbman, Raanan Fattal, and Dani Lischinski. Convolution pyramids. *ACM Trans. Graph.*, 30(6):175, 2011. 3

[5] Hany Farid and Eero P. Simoncelli. Differentiation of discrete multidimensional signals. *IEEE Transactions on Image Processing*, 13(4):496–508, 2004. 1, 2

[6] Geon-Won Lee and Jong-Ki Han. Viewport rendering algorithm with a curved surface for a wide fov in 360° images. *Applied Sciences*, 11(3):1133, 2021. 2

[7] Pavan Chennagiri Madhusudana and Rajiv Soundararajan. Subjective and objective quality assessment of stitched images for virtual reality. *IEEE Transactions on Image Processing*, 28(11):5620–5635, 2019. 1, 2

[8] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12): 4695–4708, 2012. 2, 4

[9] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 2, 4

[10] Brandy Murovec, Julia Spaniol, and Behrang Keshavarz. The role of image realism and expectation in illusory self-motion (vection) perception in younger and older adults. *Displays*, 85:102868, 2024. 1

[11] Hamid R. Sheikh and Alan C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15 (2):430–444, 2006. 4

[12] Wei Sun, Ke Gu, Guangtao Zhai, Siwei Ma, Weisi Lin, and Patrick Le Calle. Cviqd: Subjective quality evaluation of compressed virtual reality images. In *2017 IEEE international conference on image processing (ICIP)*, pages 3450–3454. IEEE, 2017. 2

[13] Yule Sun, Ang Lu, and Lu Yu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE signal processing letters*, 24(9):1408–1412, 2017. 4

[14] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006. 3

[15] Masayuki Tanaka, Ryo Kamio, and Masatoshi Okutomi. Seamless image cloning by a closed form solution of a modified poisson problem. In *SIGGRAPH Asia 2012 Posters*, pages 1–1. 2012. 3

[16] Nafiseh Jabbari Tofighi, Mohamed Hedi Elfkir, Nevrez Imamoglu, Cagri Ozcinar, Erkut Erdem, and Aykut Erdem. St360iq: No-reference omnidirectional image quality assessment with spherical vision transformers. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023. 2, 4

[17] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 1, 2, 4

[18] Xiaoyu Xiu, Yuwen He, Yan Ye, and Bharath Vishwanath. An evaluation framework for 360-degree video compression. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2017. 2

[19] Jiahua Xu, Wei Zhou, and Zhibo Chen. Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5):1724–1737, 2020. 2, 4

[20] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2):684–695, 2013. 1, 4

[21] Luyu Yang, Zhigang Tan, Zhe Huang, and Gene Cheung. A content-aware metric for stitched panoramic image quality assessment. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2487–2494, 2017. 1, 2

[22] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. 4

[23] Yufeng Zhou, Mei Yu, Hualin Ma, Hua Shao, and Gangyi Jiang. Weighted-to-spherically-uniform ssim objective quality evaluation for panoramic video. In *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pages 54–57. IEEE, 2018. 2, 4

[24] Zhe Zhu, Jiaming Lu, Minxuan Wang, Songhai Zhang, Ralph R Martin, Hantao Liu, and Shi-Min Hu. A comparative study of algorithms for realtime panoramic video blending. *IEEE Transactions on Image Processing*, 27(6):2952–2965, 2018. 3