https://doi.org/10.2352/EI.2026.38.8.IQSP-260

Applying Short-Time Fourier Transform for Flexible and Intuitive Image Quality Assessment

Subin Han, Seungwan Jeon, Yu Gyeong Lee, Sara Lee, Junho Han, DongOh Kim, KiChul Park, and Sung-Su Kim; S.LSI Division, Samsung Electronics; Hwaseong-si, Gyeonggi-do, Republic of Korea

Abstract

Evaluating image quality in natural scene images is challenging because scene composition is highly variable and image distortions often differ across regions. Existing image quality assessment (IQA) methods can quantify such distortions using frequency information, but are generally ineffective for sensor artifact detection and localized scoring. In this work, we propose a versatile short-time Fourier transform (STFT)-based framework for IQA, enabling intuitive spatial-frequency interpretation of localized patches. By performing spectral analysis within sliding windows, the STFT captures localized frequency characteristics that can be directly mapped back to their spatial positions. To improve interpretability, we incorporate region of interest (ROI)-aware patch extraction using Segment Anything 2 (SAM2) to focus the analysis on relevant areas. Within this framework, the same STFT representation can be flexibly adapted to multiple IQA scenarios through different spectral interpretations, including maze artifact detection, line-broken artifact detection, and texture scoring. Our experimental results demonstrate that the framework effectively identifies artifact regions and provides meaningful texture quality measurements; specifically, the proposed texture frequency metric achieves a Pearson correlation coefficient of 0.78 with subjective Elo scores. These results indicate that STFT-based spectral interpretation provides an intuitive and flexible approach for analyzing diverse image quality characteristics in natural scene images and supports practical workflows for image quality evaluation.

Introduction

Image quality assessment (IQA) is essential in the sensor industry for evaluating the performance of imaging devices. As the design of color filter arrays (CFAs) in the sensor pixel layer has become increasingly complicated, new image signal processing (ISP) methods for this hardware innovation are continuously being developed to produce high-quality RGB images from CFA raw data. In this context, evaluating how much these new technical advancements improve imaging performance is important for the manufacturers and their customers. They typically evaluate image quality using fundamental key performance indicators such as sharpness, noise, color accuracy, and contrast [1] , [2], and [3]. However, these indicators sometimes fail to fully address the demands of industrial product quality assurance. Particularly, sensor manufacturers often conduct IQA to analyze performance while focusing on specific modifications in their products. In this regard, sensor artifacts, which appear as periodic patterns in images, are critical concerns. Some techniques are designed to reduce these artifacts, whereas new techniques may introduce previously unseen ones. To address defects and optimize sensor registers, sensor manufacturers require flexible IQA methods to assess sensor artifacts from an industrial perspective.

Analyzing images in the frequency domain offers intuitive and informative insights for this purpose. Specifically, frequency components enable clear characterization of both image quality factors (e.g., texture and resolution) and sensor artifacts (e.g., maze pattern and line-broken artifacts). However, existing approaches often fail to fully exploit this potential. Several IQA methods have been proposed to quantify specific artifacts in high-frequency regions, such as line-broken artifacts [4] and false color artifacts [5], but they typically do not leverage frequency-domain information comprehensively. Conversely, frequency-based IQA methods typically focus on global quality [6], [7], [8]. Additionally, recent deep learning-based approaches operate as black box models without explicit use of frequency interpretation [9], [10]. While these methods are useful for assessing overall image quality, they lack localization and diagnostic capabilities required for industrial use. To address these limitations, we propose a frequency-based IQA framework that provides localized analysis suitable for industrial applications.

The short-time Fourier transform (STFT) is a well-established method for analyzing localized frequency components of signals. It provides a time-frequency representation of a signal by applying the Fourier transform to short, overlapping segments [11], allowing it to handle nonstationary signals effectively [12]. As a result, STFT has been widely used in various fields such as speech processing, biomedical signal analysis, and image processing.

In this work, we therefore propose an STFT-based pipeline for spatially resolved frequency analysis. By computing STFT on local image patches, we obtain spatially resolved spectral information that captures both periodic sensor artifacts (e.g., maze patterns and line-broken defects) and local texture characteristics. This localization capability allows defect detection as well as texture quality assessment in complex natural images. However, to extract these local spectra from semantically consistent regions across different images, we require precise region of interest (ROI) alignment regardless of viewpoint or object position. We therefore integrate Segment Anything Model 2 (SAM2) [13], a state-of-the-art segmentation model, to ensure consistent evaluation of identical objects across different images. The proposed framework supports various industrial applications ranging from artifact detection to texture scoring. Therefore, we evaluated our framework on the IQA tasks of maze artifact, line-broken artifact and even texture quality scoring to demonstrate the effectiveness and versatility of our approach.

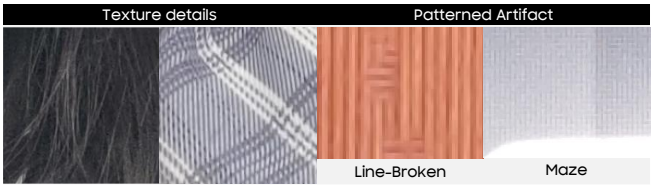


Figure 1. *Typical image regions requiring localized frequency analysis.*

Method

An overview of the proposed IQA framework is illustrated in Fig. 2. The pipeline consists of three main stages. First, spatial patches are selected from regions of interest (ROIs) to facilitate the extraction of frequency information. These patches are then transformed into the frequency domain for patch-wise analysis in the second stage. Finally, the desired final outputs are generated for individual images according to the specific objectives of the IQA task.

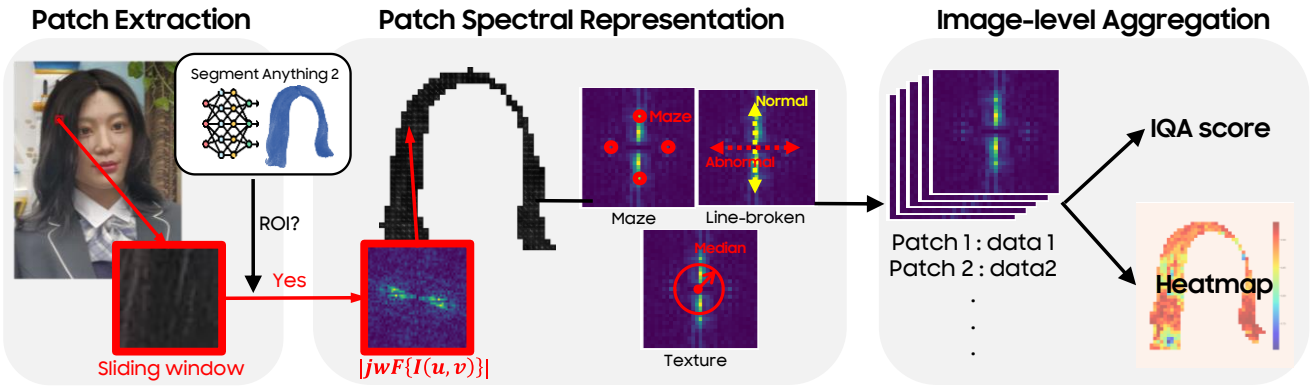


Figure 2. *Pipeline of short-time Fourier transform applied image quality assessment*

ROI-aware Patch Extraction

To focus on perceptually important regions, we first generate IQA-targeted ROI masks before sampling spatial patches. Specifically, we employ Segment Anything Model 2 (SAM2) [13], a state-of-the-art segmentation model, to identify object-level regions in the image. It is a zero-shot model capable of extracting semantic masks via prompting without requiring task-specific training. The flexibility for IQA engineers to specify bounding box, point, or mask prompts and extract segmentation masks makes this approach particularly suitable for the proposed pipeline.

Based on the generated ROI masks, we extract patches using a sliding window strategy following the conventional STFT patch sampling. This process combined traditional spatial sampling with ROI filtering: we retain only those patches where all pixels within the sliding window were fully contained within the mask. Patches exhibiting partial overlap with non-ROI regions are discarded to ensure that frequency analysis was restricted solely to the target regions.

To enable reliable frequency-domain analysis, the patch size is determined proportionally to the ROI scale to consider both spatial localization and frequency-domain representation. By adapting the patch size to the ROI dimension, the proposed approach preserved meaningful spatial context within the region while enabling stable frequency analysis of local image structures.

Patch Spectral Representation

As our goal is to capture periodic variations related to image quality assessment, we use the spatial derivatives in the frequency domain. The spectral representation using the derivative property of the Fourier transform can be efficiently computed as $|j\omega F\{I\}|$, which is obtained by multiplying STFT coefficients with frequency terms ω and j. This operation suppresses the DC component and reduces the influence of gradual shading and low-frequency illumination variations while emphasizing structural variations.

This spectral representation captures the localized frequency characteristics of each spatial patch. We interpret each patch differently depending on the specific IQA task and the results enable us to examine the signals exhibited at each original spatial position by the image. This approach allows flexible patch analysis. Spectral components can be probed, and signal frequencies can be calculated through statistical analysis. The interpretation can be adapted to target artifacts or image quality factors.

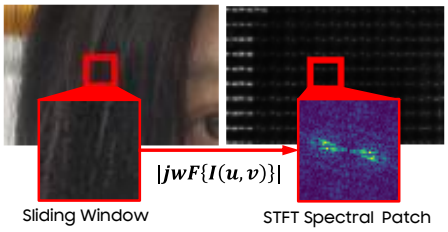


Figure 3. *Derivatives in the spectral domain.*

Image-level Aggregation

After patch-level signal analysis for the target IQA task, the results are aggregated at the image level. Patch-wise measurements can be statistically combined to produce a global image quality score for image ranking. Alternatively, patch scores can be mapped back to their spatial locations to generate artifact localization heatmaps, enabling detection and visualization of spatial quality degradations.

Experiments

Experimental Settings

Natural scene images were used for the experiments. The dataset consists of images capturing the same objects under similar physical positions using wide field-of-view (FOV) settings, while varying the imaging sensors. Specifically, sensors with resolutions of 12 MP, 50 MP, and 100 MP were used.

The experiments were conducted from three perspectives: maze artifact, line-broken artifact, and texture preservation. For artifact analysis, scenes were selected where maze and line-broken artifacts are frequently observed. For texture analysis, regions containing high-frequency patterns, such as the hair and clothing,

were used. Since maze artifacts tend to appear across the entire image, the full image was used as the ROI in this case. For the other scenarios, ROI masks were generated using SAM2 based on regions identified by an IQA engineer.

STFT-based patch extraction was performed using a sliding-window approach. The patch size was set to 5% of the ROI dimension with a 50% overlap between adjacent patches.

Maze Artifact Analysis

Maze artifacts are known to originate from color filter array (CFA) demosaicing errors in the imaging pipeline [14]. These artifacts typically appear as maze patterns in regions containing repetitive high-frequency structures. In the frequency domain, they are associated with responses around the CFA lattice frequency

$$f = \pm\frac{1}{T_{\mathrm{CFA}}}, \theta \in \{0^{\circ}, 90^{\circ}\}, \quad (1)$$

where T_{CFA} denotes the CFA period and θ represents the orientation in the frequency domain.

Based on this observation, the proposed STFT-based analysis was used to probe frequency responses around these characteristic locations. As illustrated in Fig. 4, spatial patches were first transformed into the frequency domain, and Gaussian probing was applied around the target CFA-related frequencies. Among the probed responses, the maximum value was selected as the maze signal magnitude for each patch, which was then used to construct a spatial detection heatmap.

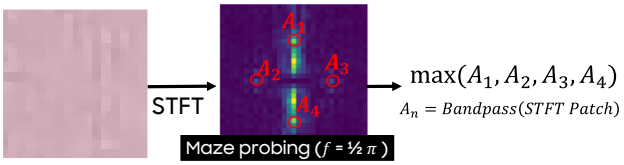


Figure 4. *Maze artifact probing and signal interpretation.*

In our experiments, we used entire image as the region of interest (ROI) in this experiment. The resulting maze responses were visualized as heatmaps to analyze the spatial distribution of the artifacts. As shown in Fig. 5, images containing strong maze artifacts exhibited larger regions with high response values, appearing as red-colored areas in the detection maps. In contrast, images without maze artifacts showed relatively low responses across the image.

A closer inspection of the cropped regions further reveals grid-like patterns that correspond to the visual appearance of maze artifacts. The corresponding cropped detection maps also exhibit strong responses in these regions, indicating that the proposed frequency probing effectively captures the spectral characteristics of maze artifacts.

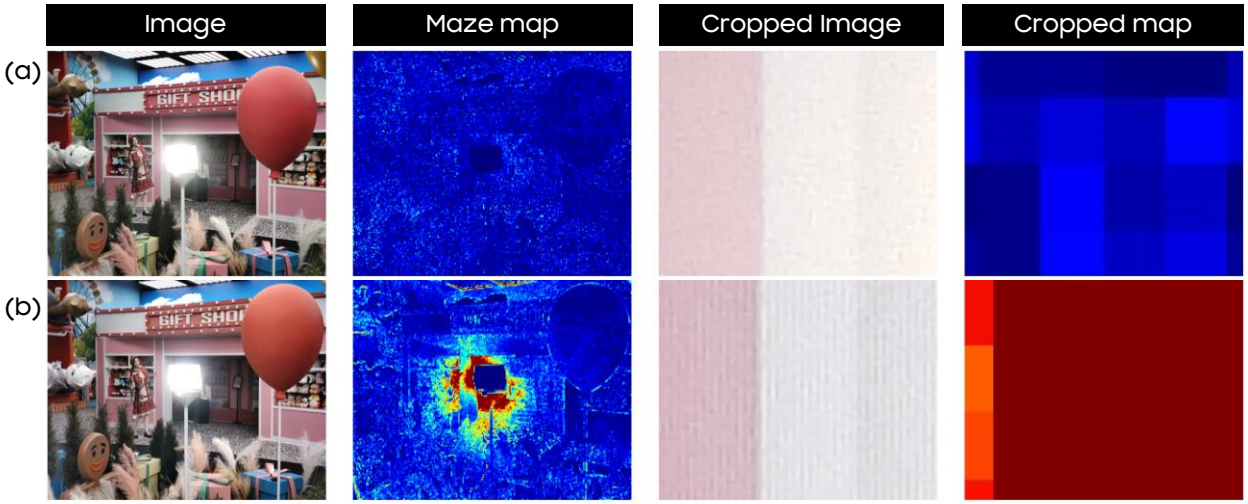


Figure 5. *Maze artifact detection results. Rows show (a) a maze-free image and (b) a maze-affected image. Columns present the original image, the maze detection heatmap, a cropped region, and the cropped detection heatmap.*

Line-broken Artifact Analysis

Line-broken artifacts are mainly caused by errors in pattern direction estimation within the image signal processor. These artifacts appear as abnormal responses along directions inconsistent with the dominant structural orientation in the image [15]. In order to observe this phenomenon, we selected ROIs with predominantly vertical patterns so that strong horizontal responses could be interpreted as line-broken artifacts.

To analyze this behavior, we first estimated the dominant signal direction from the STFT spectrum of each patch. Figure 6 shows how we measured line-broken on each STFT patches. Since the selected ROI contains structures primarily oriented in one direction, the peak response naturally indicates the main direction of the pattern. Based on this direction, we constructed an X-shaped mask spanning ±45° around the main direction to separate the signals. The two sectors aligned with the main direction were defined as the normal regions, while the opposite sectors were defined as the error regions. This process is referred to as normal-error separation, as illustrated in Fig. 6. The artifact strength in a single patch was then quantified using the anisotropic level defined as,

$$Anisotropic\ Level(dB) = -20\ \log\left(\frac{\sum Normal}{\sum Error}\right). \quad (2)$$

This metric was computed for each patch, resulting in a patch-wise anisotropic level representation.

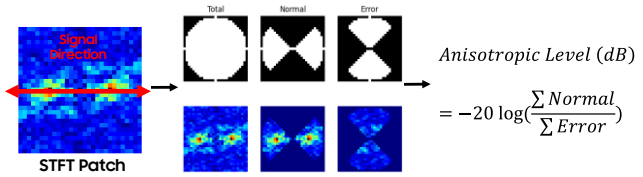


***Figure 6**. Line-broken artifact signal interpretation.*

Figure 7 shows examples of anisotropic levels computed for different patches. Patches containing fewer line-broken artifacts exhibit higher anisotropic levels, while patches with stronger artifacts show lower values. By spatially arranging these patch-wise anisotropic levels, a detection map highlights regions where line-broken artifacts are prominent. An example of the resulting detection heatmap is shown in Fig. 8.

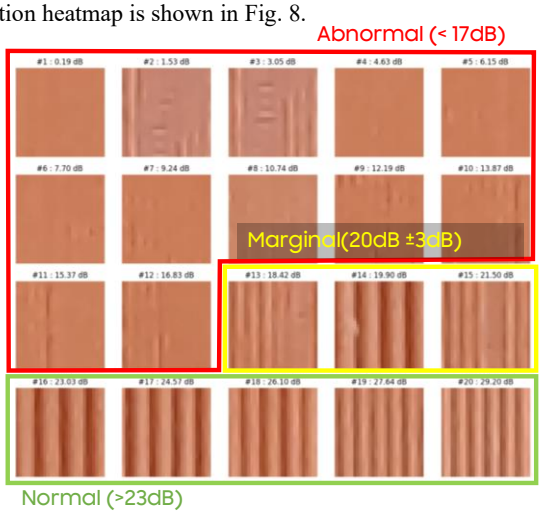


***Figure 7.** Patch-wise anisotropic level examples.*

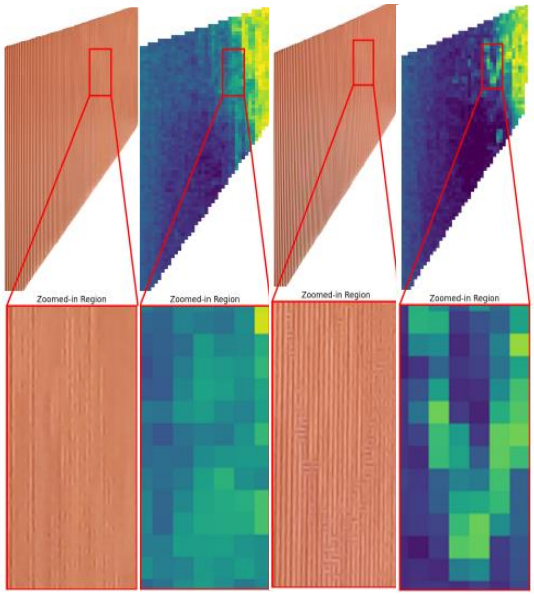


***Figure 8.** Line-broken artifact detection results.*

Texture Quality Assessment

Texture quality has been widely analyzed in the frequency domain due to its strong relationship with spatial frequency characteristics [6], [8]. To examine the applicability of our framework to this aspect, we evaluated texture quality through region of interest (ROI) frequency estimation. For this experiment, we selected ROIs containing high-frequency patterns suitable for texture assessment such as hair and textiles.

The texture analysis began with ROI-aware patch extraction followed by frequency-domain interpretation of each patch. After patch extraction, we analyzed each patch to figure out the texture frequency as Fig.9. To obtain a rotation-invariant power spectrum, polar warping was applied and the maximum response across orientations was selected for each spatial frequency, resulting in a one-dimensional rotation-invariant power spectrum. The representative texture frequency of each patch was then defined as the median frequency of this spectrum.

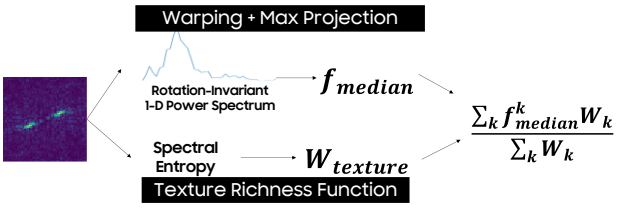


***Figure 9.** Texture frequency estimation.*

However, not all patches contribute equally to texture perception. Therefore, we introduced a texture richness weight based on the normalized spectral entropy of each patch. Frequency-domain entropy has been shown to provide discriminative information for texture analysis [16]. The normalized spectral entropy, denoted as H_{norm}, is calculated as,

$$H_{norm} = \frac{\sum_{i=1}^{N^2} p_i \log_2 p_i}{\log_2 N^2}, H_{norm} \in [0,1] \quad (3)$$

where p_i is the normalized frequency probability of the i-th frequency bin, and N is the patch size (i.e., the patch has $N \times N$ bins). The normalized entropy shows the complexity of the frequency distribution. Flat regions produce low entropy, whereas noisy regions produce high entropy. In our observations, texture-rich patches, which are important to our assessment, typically lie in an intermediate range of entropy values. As shown in Fig. 10, patches around 0.7–0.8 entropy were found to contain meaningful texture structures. To emphasize these, we defined the texture richness using a modified Tukey window:

$$R_{texture}(H_{norm}) = \begin{cases} \frac{1}{2}\left[1 + cos\left(\pi \frac{0.7 - H_{norm}}{0.05}\right)\right], & if\ 0.65 \le H_{norm} < 0.7 \\ 1, & if\ 0.7 \le H_{norm} \le 0.8 \\ \frac{1}{2}\left[1 + cos\left(\pi \frac{0.8 - H_{norm}}{0.05}\right)\right], & if\ 0.8 \le H_{norm} \le 0.85 \\ 0, & otherwis \end{cases} \quad (4)$$

where the parameters were selected. Finally, we obtained the texture frequency of the image as the weighted mean of the patch frequencies:

$$f_{texture} = \frac{\sum_k f^k_{median} W_k}{\sum_k W_k}, W_k = R_{texture}(H^k_{norm}), \quad (5)$$

where f^k_{median} is the frequency representation of k-th patch.

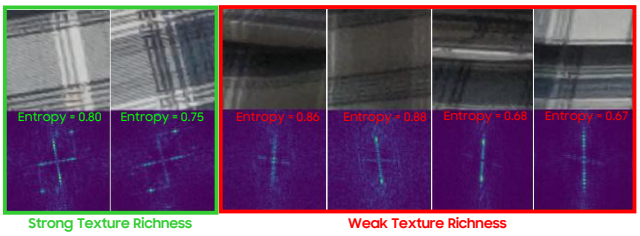


Figure 10. *Patch-wise spectral entropy distribution for texture richness estimation.*

Finally, the estimated texture frequency was used as the overall texture quality score, where higher frequencies indicate better preservation of fine texture structures. To obtain subjective references, a user survey was conducted using image patches extracted from the selected ROIs. During the survey, the ROI height was normalized across images captured with different sensor resolutions so that participants evaluated textures under comparable spatial scales. Accordingly, the texture frequency metric was expressed in the unit of line pairs per object ROI height (LP/Object ROI Height).

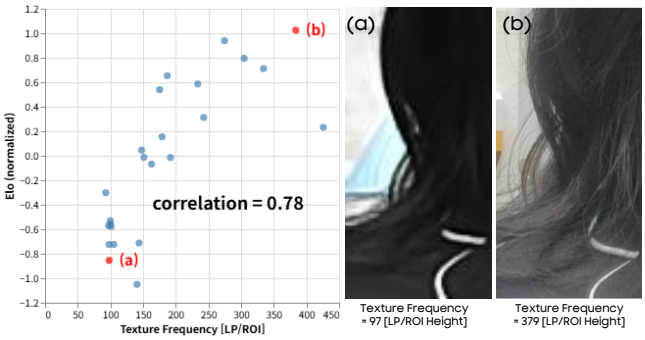


Figure 11. *Comparison of Texture Frequency and Elo Score (Subjective Image Quality Score*

Figure 11 compares the proposed texture frequency metric with the subjective Elo scores obtained from the survey. The proposed method achieved a Pearson correlation coefficient of 0.78 with the subjective scores. Although variations in object shape can occasionally produce abnormal values due to the frequency unit being defined with respect to object ROI height, the results demonstrate a meaningful level of consistency with subjective image quality assessments.

Discussion

The use of the short-time Fourier transform (STFT) representation provides several practical advantages for image quality analysis. The frequency-domain representation provides an intuitive interpretation to IQA engineers by directly examining spectral responses corresponding to specific artifacts or texture characteristics. This intuitive approach makes it straightforward to select and analyze frequency components related to particular degradations, and conversely, to quantify the degree of image quality loss. As demonstrated in Figs. 5, 7–8 and 10–11, the STFT-based analysis effectively captured maze and line-broken artifacts and provided meaningful estimates of texture frequency, illustrating the practical utility of the method. The same framework can be adapted to multiple IQA scenarios simply by applying different spectral interpretations, demonstrating its versatility and flexibility. In particular, within the sensor industry, unexpected image quality issues often need to be analyzed quickly, and the proposed framework allows engineers to investigate newly observed artifacts efficiently.

Another important aspect of the proposed approach is the use of ROI-aware analysis for natural scene images. Unlike standardized chart-based measurements, natural scene images contain objects with varying shapes and scales. Consequently, analyzing image quality at the object level provides more meaningful and reliable interpretations. By incorporating ROI segmentation prior to spectral analysis, the method directly reflects object-level structure in the frequency-domain signal interpretation. This improves interpretability, enables reliable artifact localization, and aligns with practical qualitative evaluation protocols, where IQA engineers typically select ROIs to compare perceptual differences between images.

The artifact detection heatmaps generated by the proposed pipeline also provide practical benefits for large-scale image analysis. Modern imaging sensors often exceed 100 megapixels, and manually inspecting all pixels to identify artifacts is extremely time- and effort-consuming. By highlighting regions likely to contain artifacts, the proposed method significantly reduces the manual inspection workload. Moreover, the intuitive quantification of image quality allows the derived metrics to serve as auxiliary signals in subjective surveys or as labeling guidance for data-driven quality assessment. For example, the proposed texture frequency metric achieved a Pearson correlation coefficient of 0.78 with subjective Elo scores (Fig. 11), confirming its effectiveness in quantifying perceptual quality.

Finally, as the method operates on patch-wise representations, it can be naturally integrated with modern vision architectures that process images as collections of patches, such as vision transformers (ViT). This suggests that the proposed framework could serve as a bridge between traditional signal-based image quality analysis and learning-based image quality modeling in future work, while also providing a practical tool for engineers to quickly assess and interpret image quality in natural scene applications.

Conclusion

In this work, we propose a versatile short-time Fourier transform (STFT)-based framework for image quality assessment (IQA) in natural scene images. By combining ROI-aware patch extraction with spatial-frequency analysis via STFT, the proposed framework provides intuitive and interpretable metrics for evaluating a variety of image degradations, including maze and line-broken artifacts as well as texture quality.

The proposed approach effectively captures both artifact localization and perceptual texture characteristics, as demonstrated by detection maps for maze and line-broken detection maps (Figs. 4–8), and the texture frequency analysis with high correlation to subjective Elo scores (Fig. 11, Pearson coefficient 0.78). ROI-aware spectral analysis ensures reliable evaluation in non-standardized, natural scene images, while the patch-wise representation facilitates scalable analysis of high-resolution sensors and potential integration with learning-based IQA models.

Overall, the framework offers a practical and flexible tool for IQA engineers, enabling rapid artifact assessment, quantitative assessment of perceptual quality, and generation of labeled data for data-driven modeling.

References

[1] ISO, "ISO 12233: Digital cameras — Resolution and spatial frequency responses," International Organization for Standardization, Geneva, Switzerland, 2024.

[2] ISO, "ISO 15739: Photography — Electronic still-picture imaging — Noise measurements," International Organization for Standardization, Geneva, Switzerland, 2023.

[3] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, “Subjective and objective quality assessment of image: A survey,” arXiv preprint arXiv:1406.7799, 2014.

[4] S. Jeon, Y. Lee, K. Kim, D. Yu, S.-S. Kim, and J. Yim, "Generation of reference images using filtered radon transform and truncated SVD for structural artifacts," in Proc. IS&T Int'l Symp. Electronic Imaging: Image Quality and System Performance, pp. 341-1–341-4, 2022, doi: 10.2352/EI.2022.34.9.IQSP-341.

[5] S. Han, S. Jeon, S. Lee, Y. G. Lee, H. Kim, K. Park, S.-S. Kim, and Y. Kim, "No-Reference Color Dot Artifact Assessment for Remosaiced Images," in Proc. IS&T Int'l Symp. Electronic Imaging: Image Quality and System Performance, pp. 265-1–265-6, 2024, doi: 10.2352/EI.2024.36.9.IQSP-265.

[6] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," IEEE Trans. Image Process., vol. 21, no. 8, pp. 3339–3352, Aug. 2012, doi: 10.1109/TIP.2012.2191563.

[7] Y. Liu, Z. Ni, S. Wang, H. Wang, and S. Kwong, "High dynamic range image quality assessment based on frequency disparity," IEEE Trans. Circuits Syst. Video Technol., vol. 33, no. 8, pp. 4435–4440, Aug. 2023, doi: 10.1109/TCSVT.2023.3237702.

[8] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006, doi: 10.1109/TIP.2005.859378.

[9] H. Talebi and P. Milanfar, "NIMA: Neural Image Assessment," IEEE Trans. Image Process., vol. 27, no. 8, pp. 3998–4011, 2018, doi: 10.1109/TIP.2018.2831899.

[10] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 5148–5157, 2021, doi: 10.1109/ICCV48922.2021.00510.

[11] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," Proc. IEEE, vol. 65, no. 11, pp. 1558-1564, Nov. 1977, doi: 10.1109/PROC.1977.10770.

[12] L. Cohen, Time-Frequency Analysis. Englewood Cliffs, NJ, USA: Prentice Hall, 1995.

[13] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "SAM 2: Segment Anything in Images and Videos," arXiv preprint arXiv:2408.00714, 2024, doi: 10.48550/arXiv.2408.00714.

[14] J. H. Choi, D. Y. Choi, and B. C. Song, "Demosaicking algorithm for white-RGB CFA images," IET Image Process., vol. 13, no. 5, pp. 811–816, 2019, doi: 10.1049/iet-ipr.2018.5820.

[15] Y. Kim, J. Lee, S. Kim, J. Bang, D. Hong, T. Kim, and J. Yim, "Camera Image Quality Tradeoff Processing of Image Sensor Re-mosaic using Deep Neural Network," in Proc. IS&T Int’l. Symp. on Electronic Imaging: Image Quality and System Performance XVIII, pp 206-1–206-7, 2021, doi: 10.2352/ISSN.2470-1173.2021.9.IQSP-206.

[16] M. E. Jernigan and F. D'Astous, "Entropy-based texture analysis in the spatial frequency domain.," IEEE Trans. Pattern Anal. Mach. Intell., vol. 6, no. 2, p. 237–243, 1984, doi: 10.1109/TPAMI.1984.4767507.

Author Biography

Subin Han received her B.S. degree in electrical engineering from Korea University in 2020. Since 2021, she has worked in Samsung Electronics, Republic of Korea, as an engineer. Her work has focused on computer vison, image/video signal processing, and image quality metric.

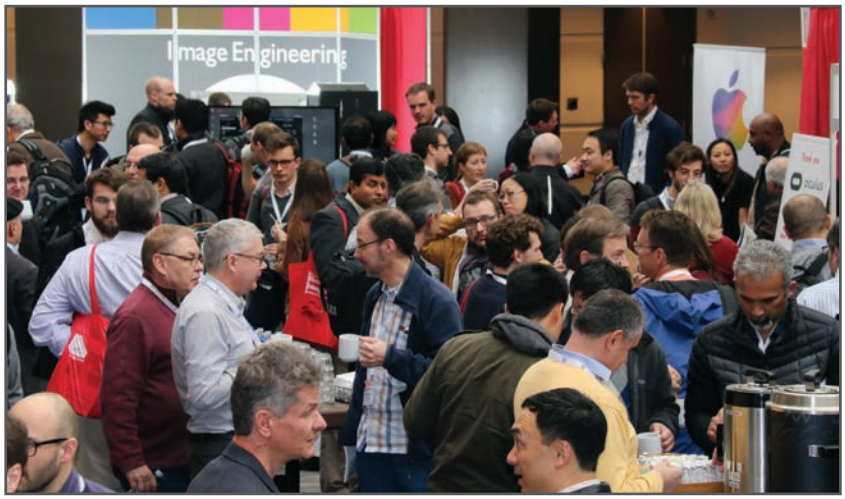
Seungwan Jeon was born in Republic of Korea in 1989. He received the B.S. degree in biomedical engineering from Yonsei University in 2014, and the Ph.D. degree in creative IT engineering from POSTECH in 2020. His Ph.D. research focused on photoacoustic/ultrasound imaging techniques using image/signal processing, beamforming, and deep learning. Since 2020, he is with Samsung Electronics, Republic of Korea, as a Staff engineer. His current research interests include camera sensor, computer vision, and image quality assessment.

Yu Gyeong Lee received her B.S. degrees in Mathematics and Computer Science from Sungkyunkwan University in 2018. Since 2019, she has worked in Samsung Electronics, Republic of Korea, as an engineer. Her work has focused on image sensors, computer vision, and image quality assessment.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org