https://doi.org/10.2352/EI.2026.38.8.IQSP-256

Depth-Aware Assessment of Spatial Frequency Response in Natural Scenes

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

Abstract

The spatial frequency response (SFR) has long been a crucial metric for evaluating imaging quality, particularly in camera performance assessment. However, the constraints of chart-based assessment limited the evaluation of natural scenes, making it challenging to evaluate resolution accurately in real-world environments. Notably, the development of the natural scene spatial frequency response (NS-SFR) has enabled resolution evaluation from natural scenes, extending its utility to diverse applications. Nevertheless, existing NS-SFR methods have been limited to two-dimensional analysis, neglecting depth-dependent behaviors such as variations in sharpness across focal planes. To address this limitation, we propose a depth-aware extension of NS-SFR, integrating depth dimension into modulation transfer function (MTF) analysis, and establish a model of the depth-MTF relationship that derives a representative MTF value for a single image's resolution. Our approach extends conventional planar NS-SFR analysis into a 3D depth-augmented framework that accounts for depth-dependent variations in MTF. Also our results suggest that our approach enables a more resilient and informative methodology for accurate cross-sensor comparison, yielding predictions that show a reasonable correspondence with resolution tendencies observed in natural scenes, while enhancing robustness under varying illumination.

Introduction

Image sharpness is a critical factor in evaluating camera performance and assessing overall imaging quality. It plays a key role in camera performance because it quantifies how well a camera depicts the details. The spatial frequency response (SFR) has long been used to quantify resolution characteristics in imaging systems. Especially, the edge-based Spatial Frequency Response (e-SFR) [1] based on the ISO 12233 standard is used to measure sharpness. In e-SFR analysis, which analyzes slanted edges on a test chart, once the slanted edge is identified, the pixels are projected along its direction to extract an edge profile, profile This edge profile is then used in this process (pointing to the material) to extract the modulation transfer function (MTF) [2] which represents the image's sharpness. While chart-based methods provide controlled evaluation environments, they exhibit significant limitations when applied to natural scenes, hindering accurate assessment of resolution in real-world environments.

However, conventional chart-based SFR measurements are limited to controlled laboratory conditions and cannot fully represent the imaging performance encountered in real-world scenes. To address these limitations, the Natural Scene Spatial Frequency Response (NS-SFR) [3] was developed to extend sharpness estimation to real-world images. Unlike standard e-SFR, which requires fixed, predetermined chart edges, NS-SFR introduces algorithms to detect and extract suitable step edges directly from natural scene images by detecting and analyzing step edges present in the scene. Development in NS-SFR enabled resolution assessment in diverse applications such as automotive sensors, computer vision, and immersive technologies like AR/VR.

However, although NS-SFR enabled resolution assessment under realistic imaging conditions, existing approaches remain restricted to two-dimensional analysis and thus fail to account for depth-dependent behaviors—such as sharpness variations across focal planes or Depth of Field (DoF) effects— which are inherent in real-world capture. Consequently, depth-dependent imaging characteristics are not explicitly analyzed. Moreover, existing NS-SFR approaches remain susceptible to variations in edge angle and noise levels [4-6], further compromising reliability when comparing images from different sensors that may exhibit significantly different out-of-focus regions or depth distributions.

Fortunately, in parallel with NS-SFR development, advances in monocular depth estimation [7-9], which provide the opportunity to infer depth information directly from single images have demonstrated that natural scenes contain rich three-dimensional information, thereby enabling both the interpretation of depth information and the incorporation of related characteristics, such as spatially varying sharpness that conventional metrics fail to capture.

Building upon recent advances in depth estimation from a single image, this paper proposes a depth-aware extension of the NS-SFR framework that incorporates aforementioned depth estimation into the MTF analysis while addressing the aforementioned limitations of existing NS-SFR. By integrating depth information with spatial frequency analysis, it becomes possible to examine imaging performance across the depth dimension and better characterize depth-dependent resolution behavior, by expanding conventional planar MTF analysis to a three-dimensional formulation, thereby providing a more comprehensive evaluation of imaging performance in natural scenes.

The paper is organized as follows. We begin by detailing the methodology for establishing the depth-MTF functional relationship, drawing upon the aforementioned NS-SFR backgrounds, thereby enabling explicit characterization of depth-varying sharpness and enhanced measurement robustness. Subsequently, to evaluate the proposed approach, we investigate the influence of depth information on MTF performance and assess correlations between edge factors and MTF estimates, with validation conducted on real-world natural images across heterogeneous sensors and various illumination conditions. Finally, we examine the implications of the proposed method and subsequently outline future research directions for further advancement of NS-SFR.

Method

The proposed depth-aware NS-SFR framework is summarized in Figure 1, consisting of three main stages: (1) valid edge-based MTF and depth measurement, (2) depth–MTF association, and (3) depth-weighted average MTF calculation. The detailed methodology is presented below.

Valid Edge-based MTF and Depth Measurement

Extraction of Valid Edges

Building upon the conventional NS-SFR framework proposed by Van Zwanenberg [3], we firstly extract candidate slanted edges across the entire input natural image. These edges are isolated by extracting candidate Regions of Interest (ROIs) centered on the detected edges. Subsequently, the candidate ROIs are validated based on multiple criteria [3], including edge angle, contrast, ROI size, linearity, and edge transition behavior to ensure that the edges are suitable as step edges.

MTF Calculation

From each validated edge ROI, ESF is obtained by projecting pixel values along the direction perpendicular to the edge. These ESFs are then used to derive SFR and corresponding MTF values, such as MTF50, MTF25, and MTF10. These values at specific edge locations serve as each point data describing MTF distribution across the depth dimension. Specifically, in this study, we employ the detailed processing procedures and parameters of obtaining MTF values detailed in [6].

Depth Estimation

To associate depth information with the extracted edge measurements, a monocular depth estimation model is applied to the input image. In this study, we used Depth-Anything-V2-Large [7], which is one of the monocular depth estimation model to generate a 3D depth map from a single RGB image. We adopt the large variant of the model with approximately 335 million parameters, to balance between accuracy at object boundaries or edges where we are interested in, particularly in regions of complex scene structures and efficient runtime. Also, the input image was resized to 518 × 518 pixels during inference.

The generated depth map provides a relative depth estimated value for each pixel in the image. For each edge ROI, the representative depth value is determined using the top 10% of depth values within the dilated edge mask region, under the assumption that the edge to be examined is the region closest to the sensor.

Depth–MTF Association

Assuming a correlation between the sensor's depth of field (DoF) and MTF values (specifically MTF50, MTF25, and MTF10), this study proposes a mathematical model to describe the relationship between depth and sharpness. Firstly, once depth values are assigned to each edge ROI, the corresponding MTF values are paired with their estimated depth values. This produces a dataset consisting of:

$$\{(depth_i,\ MTF_i)\}_{i=1}^{N}$$

denoting the set of observed data points, where N is the total number of samples. Here, x_i and y_i represent the i-th samples of depth value and the corresponding MTF value, respectively.

Depth-Aware MTF Modeling

To facilitate modeling, we employ a Gaussian Beam Propagation model [10], which characterizes the Gaussian distribution representing the depth-dependent sharpness profile. Specifically, radius of beam, denoted by $w(z)$ is assumed to follow a square root of a quadratic function, as given in Equation (1).

$$w(z) = w_0\sqrt{1+\left(\frac{z}{z_R}\right)^2} \tag{1}$$

where w_0 is the beam waist, z is the distance propagated from the plane and z_R is the Rayleigh range. In our model we assumed that beam waist and MTF is inversely proportional, despite the underlying Fourier transform and analytically complex non-linear operations. Accordingly, the relationship between distance and MTF is modeled using a parametric function

$$f(d;\theta) = \frac{a}{\sqrt{\left(1+b\cdot(d-c)\right)^2}} \tag{2}$$

,where $\theta = \{a, b, c\}$ are fitting parameters and d to be the relative depth value. The parameters are estimated by minimizing a non-linear least squares objective defined as

$$\min \sum_{i=1}^{N}\left(MTF_i - f(d_i;\theta)\right)^2 \tag{3}$$

Due to the non-linear form of the model, the optimization is performed using an iterative non-linear least squares solver employing the Trust Region Reflective (TRF) algorithm implemented in the SciPy library [11]. To ensure robustness against outlier MTF, the values obtained directly from the fitted depth-MTF model are not used as-is. Instead, we use the model-predicted MTF values as a weighting factor- hereafter referred to as the depth weight— which is applied to each ESF. Further details on this weighting strategy are provided in the subsequent section.

Average MTF calculation

In the final stage, individual ESFs are aggregated to derive a scene-representative Average SFR, following the aggregation scheme described in [6]. Rather than employing simple averaging, each ESF is weighted by a depth weight obtained from the model-predicted depth-MTF relationship with the weighting procedure following [12], thereby enhancing robustness against MTF outliers and noise variability. The resulting weighted average ESF is then used to compute the final average SFR, from which a scene-representative MTF value is subsequently derived. This enables more accurate assessment of the sensor's imaging quality and facilitates reliable performance comparison across heterogeneous sensors.

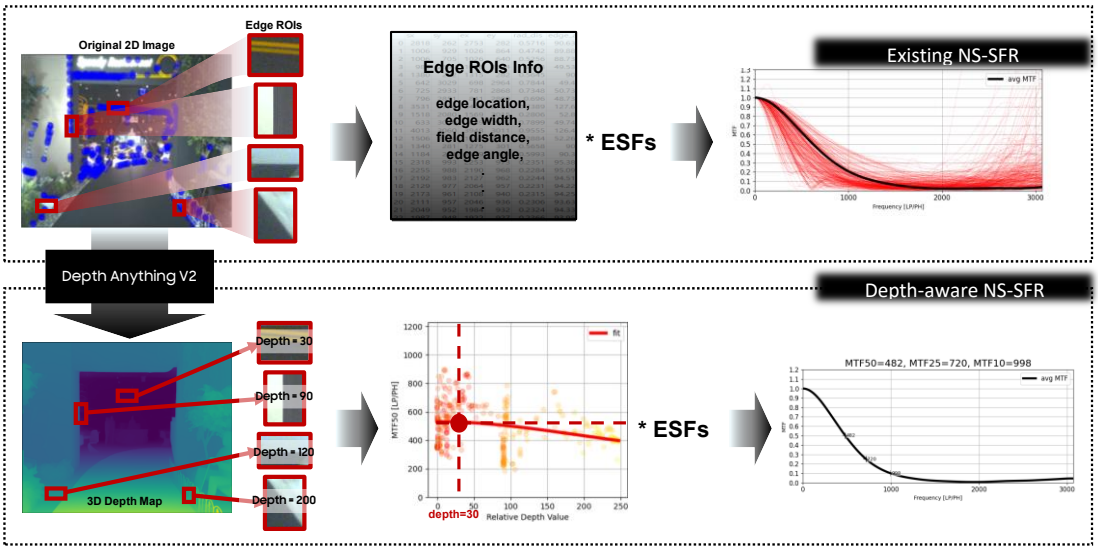


Figure 1. *Comparison of flowchart between the existing NS-SFR method and the proposed depth-aware method. The existing NS-SFR method (top row) basically follows the standard NS-SFR procedure, and the average MTF is obtained by weighting the SFR graphs of each edge with weight factors computed from numerous edge ROI attributes, such as edge location edge width, field distance, edge angle, and so on. The proposed method (bottom row) formulates the MTF vs. Depth relationship and derives the depth weight from the estimated function orange line), which is consequently used as weight factor when computing the average MTF*

Results

To begin with, we evaluated the influence of the depth factor on MTF value, specifically MTF50, due to its relative robustness to noise and artifacts. Subsequently, the effectiveness of the proposed method is demonstrated on natural scenes, validating its capability across sensors and scene conditions.

Assessment of Correlation among Edge Factors

Among conventional edge-related factors [12] including field distance, chief ray angle (CRA), and edge angle, the influence of the angle factor on MTF was investigated in various natural scene and compared with the depth factor in this work. Using scenes exhibiting substantial depth variation, the correlation between measured MTF and each factor was quantitatively analyzed. As illustrated in Figure 2, the depth factor (Fig. 2b) demonstrates a significantly stronger correlation with MTF than the angle factor (Fig. 2a), indicating that depth serves as a dominant factor of image sharpness in natural scenes. Furthermore, this result suggests that depth-axis analysis can provide a more robust and reliable framework for assessing imaging performance under complex, real-world conditions.

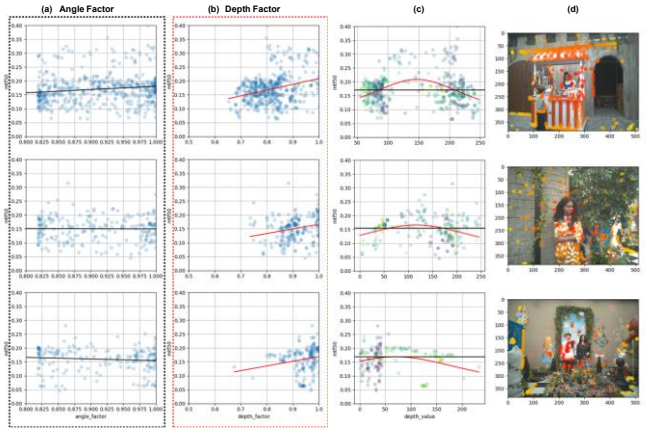


Figure 2. *Edge-factor–MTF analysis for three representative scenes. All images have a resolution of 12 MP (a) MTF vs. angle factor – blue dots denote MTF50 values for each edge; black line indicates the fitted linear trend. (b) MTF vs. depth factor – blue dots denote MTF50 values for each edge; red line indicates the fitted linear trend. (c) MTF vs. relative depth – colored dots denote MTF values at their respective depths; red curve denotes the fitted non-linear depth–MTF model. (d) Representative scenes with colored dots marking each edge location.*

Test on Real Scene with Cross-Sensor Comparison

The proposed method was validated on real natural images, demonstrating that the estimated DoF effectively captures the overall trend of depth-dependent MTF (Figure 3), which follows an inversely proportional Gaussian beam model. Furthermore, the method enables comparative analysis across sensors by analyzing MTF distributions along the depth axis using ROIs at identical spatial locations across varying depths. As shown in Figure 3, pairwise comparison between Sensor #1 and Sensor #2 which is presented in the top and middle rows, respectively, shows that the

predicted depth–MTF profiles in the bottom row are well aligned with the spatial-resolution characteristics observed in real image patches of different five depth regions (a-e), confirming strong agreement between the fitted depth–MTF model and actual observations. Notably, while conventional MTF50 indicates only a minor difference between the sensors (approximately 13 LP/PH), the proposed depth-aware analysis reveals a more discriminative and physically meaningful performance evaluation. Specifically, Sensor #1 exhibits superior MTF performance at relative depths below approximately 80 (indicated by the vertical line in Figure 3), consistent with the image sharpness observed in the corresponding regions. Conversely, beyond this threshold, the performance ranking hierarchy reverses, with Sensor #2 outperforms Sensor #1. This crossover behavior is consistently reproduced in both the predicted profiles and real image observations, demonstrating the reliability and effectiveness of the proposed framework for multi-dimensional, depth-aware imaging performance evaluation.

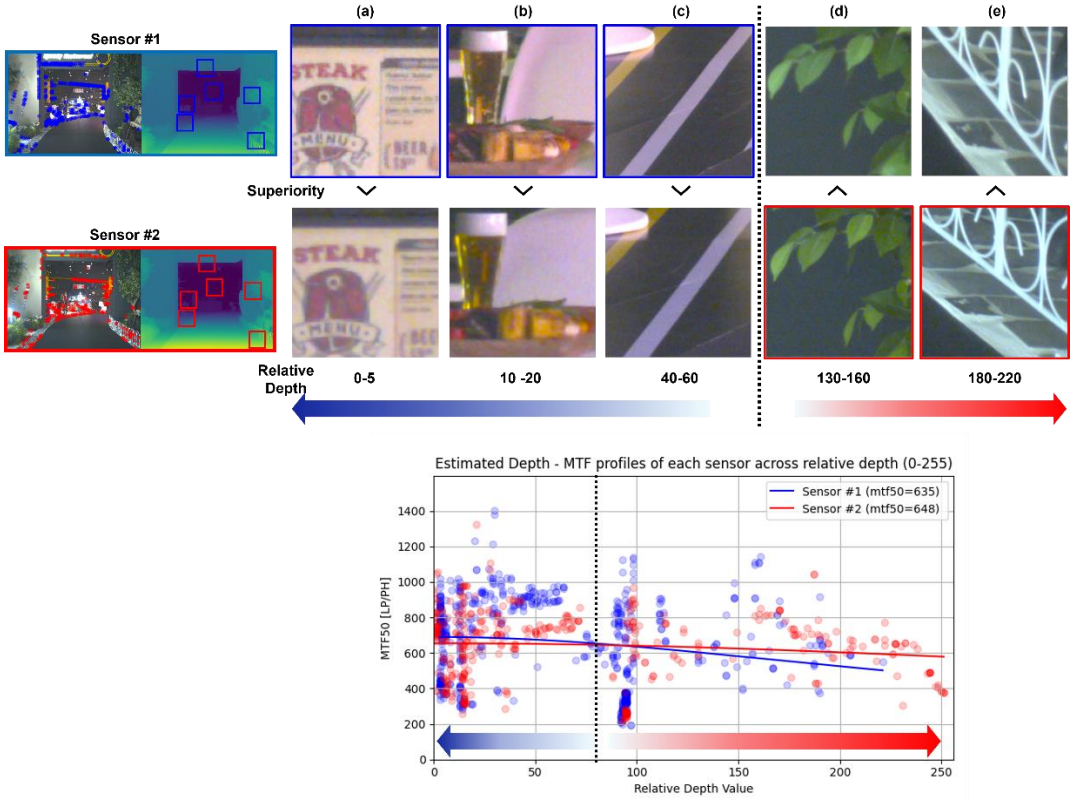


Figure 3. *Comparative depth -MTF analysis for the heterogeneous sensors (Sensor #1 = blue, top row; Sensor #2 = red, middle row). In the top and middle rows, the left figure shows the valid slanted edge locations and their corresponding positions on the depth map, while the five real-scene image patches on the right, labeled (a)–(e), covering five distinct depth ranges. The bottom row shows the relative depth -MTF plots obtained from all valid edges and the corresponding depth-MTF profiles. The vertical black dashed line indicates a relative depth value of ≈ 80, where the performance superiority switches: Sensor #1 yields higher MTF for depths <80, while Sensor #2 outperforms for depths > 80. The same blue and red color coding is used for both the image patches and the MTF curves throughout the figure. Additionally, images from Sensor #1 and Sensor #2 have a resolution of 12 MP.*

Actual Test on Varying Scene Conditions

Although prior NS-SFR approaches improve noise robustness, residual deviations remain under varying illumination conditions [4-6]. With the application of the depth-weighting factor, a reduction in illumination-dependent MTF variation is observed, leading to improved robustness across different lighting conditions.

In this study, the effectiveness of the method was evaluated using nine real natural-scenes across a wide range of lighting conditions (Indoor light only, 1 lux, 5 lux, 20 lux, 1,000 lux, and 30,000 lux), with an average of 5.5 illumination levels analyzed per scene. A brief description of each scene is provided in Table 1. As shown in Figure 4, the proposed approach consistently reduces the standard deviation (STD) of MTF measurements across most scenes, achieving an average reduction of 45.1% compared to the conventional method [6]. This improvement demonstrates enhanced robustness against illumination-induced noise variability and ensures more stable and consistent MTF estimation, thereby providing a reliable quantitative basis for image quality assessment even under challenging low-light and noisy condition.

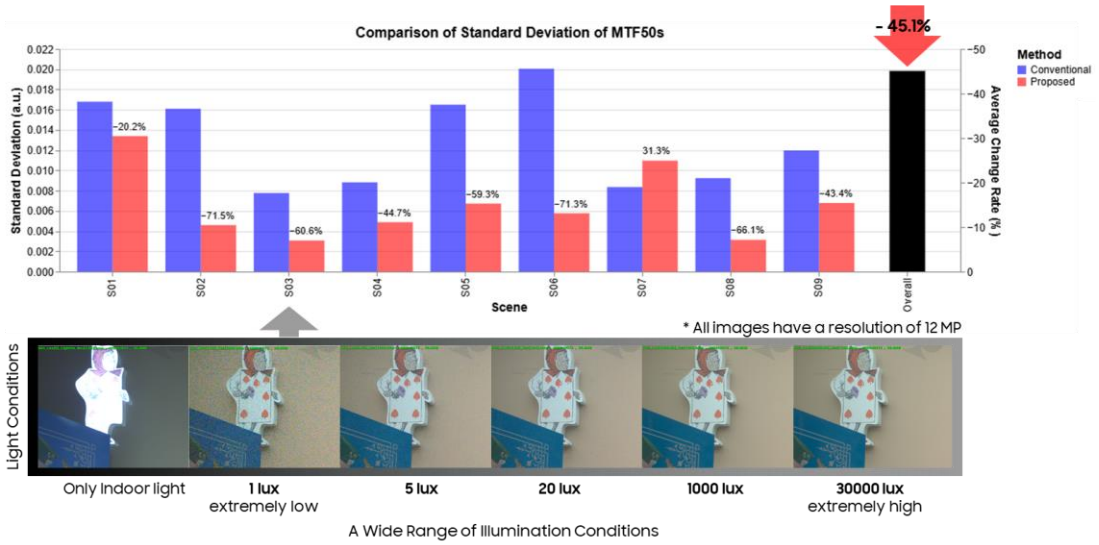


Figure 4. *Comparison of the standard deviation (STD) of MTF values across different lux conditions across nine scenes before and after applying the proposed method. (a) Comparison of STD values for scenes S01–S09 obtained using the conventional method and the proposed method. The rightmost "Overall" represents the average STD reduction ratio across all nine scenes. The STD decreased in most scenes with the proposed method, indicating improved measurement stability under varying illumination conditions. The only exception is S07, where the STD increased slightly due to multiple bright light sources in the background, which introduced additional measurement variability (b) Example scene showing the range of illumination conditions used for STD evaluation within a single scene. Additionally, all images have a resolution of 12 MP.*

Table1. Abstracted Description of Each Scene in Figure 4

Scene Number	Description
S01	big vendor stand, deeply shadowed space on the side
S02	standing woman, street, shallow depth
S03	two people, flowers on the front, mid depth
S04	seated person, deep depth
S05	checkout counter, dark background, indoor
S06	booth, background lights, shallow depth
S07	indoor, table, shallow depth, strong background lights
S08	store, convergence toward the center
S09	Shop, menus, shallow depth

Conclusion

In conclusion, this work extends conventional NS-SFR analysis from planar MTF evaluation to a 3D depth-augmented framework, introducing the novel functional modeling of the depth-MTF relationship. Through this approach, we demonstrate that depth exhibits a stronger correlation with MTF, establishing it as a critical factor in resolution analysis. Experimental results demonstrate that the proposed approach enables more robust resolution estimation and facilitates reliable cross-sensor comparison under realistic imaging conditions.

These advancements enhance the practical applicability of NS-SFR, enabling more precise and reliable real-world image quality evaluation through a more comprehensive assessment of imaging quality. Furthermore, the framework provides a foundation for depth-aware NS-SFR, contributing to improved performance in applications such as smartphone imaging, autonomous driving, and AR/VR systems, as well as a more quantitative basis for visual fidelity.

Future work will focus on improving the depth–MTF fitting by employing higher-accuracy and metric depth estimation models, particularly with tele-sensor data for enhanced far-distance analysis. Incorporating lens parameters, additional edge-related factors, and robust fitting techniques (e.g., Huber, RANSAC) is also expected to further enhance the accuracy and reliability of the proposed NS-SFR framework.

References

[1] British Standard Institute, in BS ISO 12233:2017 Photography - Electronic still picture imaging – Resolution and spatial frequency responses, BSI Stand. Publ., 2017, p. 1–62.

[2] J. Yu Zhang Y, Qi B, Bai X, Wu W, Liu H. "Analysis of the Slanted-Edge Measurement Method for the Modulation Transfer Function of Remote Sensing Cameras". Applied Sciences. vol. 13, no. 24, Art. no. 13191, 2023.

[3] O. van Zwanenberg, et al. "Estimation of ISO12233 Edge Spatial Frequency Response from Natural Scene Derived Step-Edge Data," Journal of Imaging Science and Technology (JIST), vol. 65, no. 6, pp. 60402-1-60402-16(16), 2021.

[4] O. van Zwanenberg, et al. "Natural Scene Derived Camera Edge Spatial Frequency Response for Autonomous Vision Systems," in IS&T/IoP London Imaging Meeting, London, 2021.

[5] O. van Zwanenberg, et al. "Analysis of natural scene derived spatial frequency responses for estimating camera ISO12233 slanted-edge performance," in Journal of Imaging Science and Technology, vol.65, pp. 060405-1 - 060405-12, 2021.

[6] S. Lee, et al. "Angle-and Noise-robust Spatial Frequency Response-based Resolution Analysis in Natural Scenes." in IS&T International

Symposium on Electronic Imaging: Electronic Imaging, San Francisco, 2025.

[7] L. Yang, et al. "Depth Anything v2."Advances in Neural Information Processing Systems 37 (NeurIPS), 2024.

[8] D. Eigen, C. Puhrsch, and R. Fergus, "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network." Advances in Neural Information Processing Systems 27 (NeurIPS), 2014.

[9] J. Zhang, et al. "Towards Robust Monocular Depth Estimation in Non-Lambertian Surfaces." European Conference on Computer Vision (ECCV), Springer Nature Switzerland, 2024.

[10] Edmund Optics, "Gaussian Beam Propagation," Edmund Optics Knowledge Center. [Online]. Available: https://www.edmundoptics.com/knowledge-center/application-notes/lasers/gaussian-beam-propagation/ [Accessed: Nov. 12, 2025].

[11] P. Virtanen, et al., “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” Nature Methods, vol. 17, pp. 261–272, 2020.

[12] O. Van Zwanenberg, "Camera Spatial Frequency Response Derived from Pictorial Natural Scenes," Diss, University of Westminster, School of Computer Science and Engineering, 2022.

Author Biography

Sara Lee received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, in 2019, and 2021, respectively. Her M.S. work focused on ultrasound imaging and signal processing. Since then she has worked in Samsung Electronics, Republic of Korea, as an engineer. Her work has focused on image sensor, image/video signal processing, computer vison, and image quality assessment.

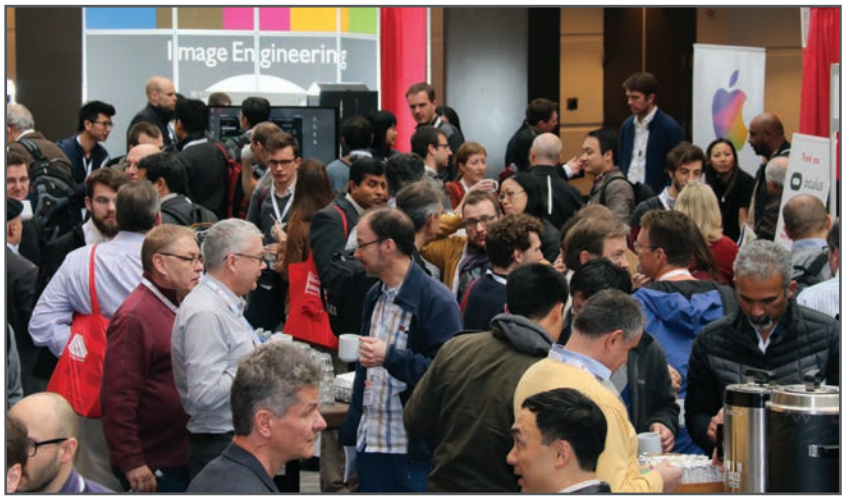
Yu Gyeong Lee received her B.S. degrees in Mathematics and Computer Science from Sungkyunkwan University in 2018. Since 2019, she has worked in Samsung Electronics, Republic of Korea, as an engineer. Her work has focused on image sensors, computer vision, and image quality assessment.

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.

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