Natural Scene Derived Camera Edge Spatial Frequency Response for Autonomous Vision Systems

Van Zwanenberg, O., Triantaphillidou, S., Jenkin, R. and Psarrou, A.

This is an electronic version of a paper presented at Van Zwanenberg, O., Triantaphillidou, S., Jenkin, R. and Psarrou, A. 2021. Natural Scene Derived Camera Edge Spatial Frequency Response for Autonomous Vision Systems. IS&T/IoP London Imaging Meeting 2021. London 20 - 22 Sep 2021 The Society of Imaging Science and Technology. .

Natural Scene Derived Camera Edge Spatial Frequency Response for Autonomous Vision Systems

Oliver van Zwanenberg, Sophie Triantaphillidou, Robin B. Jenkin, and Alexandra Psarrou
School of Computer Science and Engineering, University of Westminster, UK

Abstract

The edge-based Spatial Frequency Response (e-SFR) is an established measure for camera system quality performance, traditionally measured under laboratory conditions. With the increasing use of Deep Neural Networks (DNNs) in autonomous vision systems, the input signal quality becomes crucial for optimal operation. This paper proposes a method to estimate the system e-SFR from pictorial natural scene derived SFRs (NS-SFRs) as previously presented, laying the foundation for adapting the traditional method to a real-time measure.

In this study, the NS-SFR input parameter variations are first investigated to establish suitable ranges that give a stable estimate. Using the NS-SFR framework with the established parameter ranges, the system e-SFR, as per ISO 12233, is estimated. Initial validation of results is obtained from implementing the measuring framework with images from a linear and a non-linear camera system. For the linear system, results closely approximate the ISO 12233 e-SFR measurement. Non-linear system measurements exhibit scene-dependant characteristics expected from edge-based methods. The requirements to implement this method in real-time for autonomous systems are then discussed.

Introduction

Deep neural networks (DNNs) are currently used as one of the main technologies in image recognition tasks. However, recent studies have shown that DNNs are susceptible to adversarial natural noise in input images [1, 2] that degrade their performance and sometimes result in unpredictable object or scene classification decisions. With the application of DNNs in decision critical systems, such as autonomous vision systems, it is important to develop camera performance measures that can monitor the output quality at any given moment, i.e., in real-time.

There are several reasons why camera signal may deteriorate during real-time operation, including camera system failure, motion blur, defocus, and environmental conditions. Through live monitoring, the DNN can be adapted according to the measured and expected camera performance, adjusting the image signal processing (ISP), or completely removing the automation when the SFR drops under what is deemed safe operation.

The ISO12233 e-SFR is a standardised method for measuring camera system performance from slanted edges [3]. The SFR is an adaptation of the Modulation Transfer Function (MTF). MTFs/SFRs are traditionally obtained from captured test charts under laboratory conditions. This study attempts to estimate the traditional e-SFR measurement using the natural scene derived SFR (NS-SFR) measures, as previously presented in [4, 5]. Thus, laying the foundation for monitoring live camera performance whilst providing results that are in accord with an accepted and much-used standard.

Measuring camera performance obtained directly from natural scenes is not a new concept. In several applications, step-edges are selected from images for camera system characterisation. Such applications include assessing aerial camera systems [6] and optimising digital scan resolution for film archives [7]. The texture-MTF method [8] was recently modified to work with input images of natural scenes for evaluating scene dependency of non-linear camera systems [9, 10]. This technique works effectively with little computation, but it assumes that the input noise power spectrum is known, making it unsuitable for real-time measurements.

To date, the most effective approach to measure camera performance from natural scene images is a prediction of the Point Spread Function (PSF) through a convolutional neural network (CNN) to then calculate the MTF [11]. This approach yields accurate MTF estimates for linear systems, with computational times of a few minutes per image. However, out of focus or textureless image regions produce errors in the estimated MTF, or return non-predictions. This method is not based on a traditional technique, and therefore, conventional SFR/MTF measures cannot be compared with its estimates.

Background: The NS-SFR

Previous work proposed a novel framework to adapt the ISO 12233 e-SFR to measure the NS-SFR [4, 5]. This methodology is based on an automated process that selects step-edge regions of interest (ROIs) from natural scene images.

The NS-SFR methodology extracts, isolates, and validates suitable step-edges from natural scene images to derive the camera SFR by subsequently implementing the standard e-SFR (slanted-edge) algorithm. Unlike e-SFRs obtained from test charts with known edge contents, the NS-SFRs are derived from step-edges with unknown spatial frequency content. Thus, the measure inherently contains variations since it accounts for both camera performance and scene contents. Its output cannot be classed as an e-SFR; it is therefore referred to as an NS-SFR.

NS-SFRs for a given camera system and setting form an envelope. Such envelopes were shown to be scene-dependant due to selected edge locations, surrounding scene texture (noise) and depth of focus/depth of field.

Further, variations in the SFR parameters, such as edge angle, edge contrast and the ROI size, are also shown to introduce variation in the measured NS- SFRs [4, 5]. This study takes the NS-SFR measure a step further by examining each parameter's range, thus, analysing and regulating the sources of such variations to allow for a stable estimate. These 'calibrated' NS-SFRs are then used to estimate the system e-SFR, i.e., a measure designed to match the e-SFR obtained using the ISO 12233 test chart. The system e-SFR estimation is validated using two camera systems, one with linear performance and the other incorporating highly non-linear ISP. The paper finally briefly discusses the advantages and caveats of the proposed measure and requirements for real-time implementation.

Parameter Range for e-SFR estimation

The NS-SFR data is derived from isolated edges in ROIs with a range of parameters. These include edge angle, edge contrast, and ROI height and width. The first step in deriving the system e-SFR estimate was to reduce extensive NS-SFR variations by determining suitable parameter ranges. This step

should be applied without restricting the amount of valuable NS-SFR data, as suitable edges are not commonly found in natural scenes.

A large range of step-edge angles and contrasts, region of interest (ROI) heights and widths, and signal-to-noise ratios (SNRs) were tested, using edges captured from standard test charts and camera simulations. In addition, edge isolation techniques applied in the NS-SFR methodology, such as pixel stretching [5], were applied to the selected ROIs to evaluate their impacts on the system e-SFR variation. The results gave details and information that previous e-SFR variation evaluations and benchmarking publications had not [12–14].

Figures 1 and 2 demonstrate some of the findings. Results in both figures are from simulations. Figure 1 used ROIs with image noise level set to SNR18 and utilised the mean absolute error (MAE) from the ISO12233 e-SFR (SFR measured from a noiseless ROI with the standard e-SFR parameters) to colour map the scatter plot data. The MAEs were calculated from spatial frequencies 0 to 0.5 cyc/pixel. SNR18 was used to introduce a high noise level to illustrate a high MAE variation introduced by these parameters. The image noise was simulated using both Poisson and Gaussian distributions, corresponding to shot and read noise. Figure 2 shows the edge angle MAE introduced in the higher frequencies (0.4 to 0.5 cyc/pixel), with and without pixel stretching. The figure illustrates how this isolation method reduces the effects of image noise on the measured SFR, but the variation caused by angle changes remains. The noise reduction allows the possible SFR parameter range to be expanded. The other SFR parameters examined in this study, i.e., edge contrast, ROI height and width, showed similar trends.

Using SFR variance data, the NS-SFR parameter ranges adequate for estimating the system e-SFR were established. All parameter values are listed and compared to the recommended ISO 12233 standard parameters in Table 1.

Parameter	ISO 12233 e-SFR	NS-SFR
ROI Size	>64 x 80-500 pixels	>20 x 20-130 pixels
Edge Angle	<45° (5° Recommended)	2.5° – 35°
Edge Contrast	0.55 – 0.65 Michelson Contrast	0.55 – 0.65 Michelson Contrast

Table 1. The parameter ranges for the e-SFR ISO 12233 method against those used to estimate the system-SFR from the NS-SFR data.

ROI with isolated step-edges from natural scenes need ideally to be small to reduce the probability of including unwanted artifacts, such as changes in illumination and focus across the edge, double-edges and overlapping scene structures. However, small ROIs have been shown to introduce higher error due to image noise and insufficient edge data points. Pixel stretching reduces this error as the effects of noise were reduced, allowing small ROI heights of a minimum of 20 pixels and widths of 20 pixels, as long as the ROI neither interferes with the edge nor inhibits the Edge Spread Function (ESF). Although smaller ROIs are shown to be usable, if there are larger ROIs available in the selection process [5] they are prioritised.

It is well documented that the smaller the edge angle from the vertical, the less error is introduced in the measured SFR [12–14]. However, significantly restricting the angle limits the number of edges isolated from natural scenes. Edge angle, therefore, was kept within a broad range, 2.5 to 35 degrees.

In the working examples, contrast variations did not introduce a large error. Nonetheless, contrast provokes non-linear image processing changes, so it was kept in a narrow range to minimise non-linear ISP effects, as recommended in the ISO 12233.

System e-SFR Estimation Methodology

With the SFR parameter ranges established, the system e-SFR can be estimated via the following four steps:

1. To minimise optical imaging circle performance variation, the frame is segmented into uniform radial distances. The number of segments is adjustable for different application requirements. In this study, six radial distances were used.
2. For each radial distance, the distribution of the ROI Line Spread Function (LSF) half peak widths is analysed. Isolating the narrowest LSFs for the system e-SFR estimation, i.e., the edges most likely to be a response from a perfect step-edge input. This study uses the 10th percentile of the LSF half peak width distribution for this purpose.
3. The selected ROIs, per radial distance, are assigned to a multi-dimensional grid, binning the NS-SFRs to represent the output with unique combinations of SFR parameters. This binning process helps to reduce any anomalous NS-SFR values and bias due to larger quantities of specific parameters.
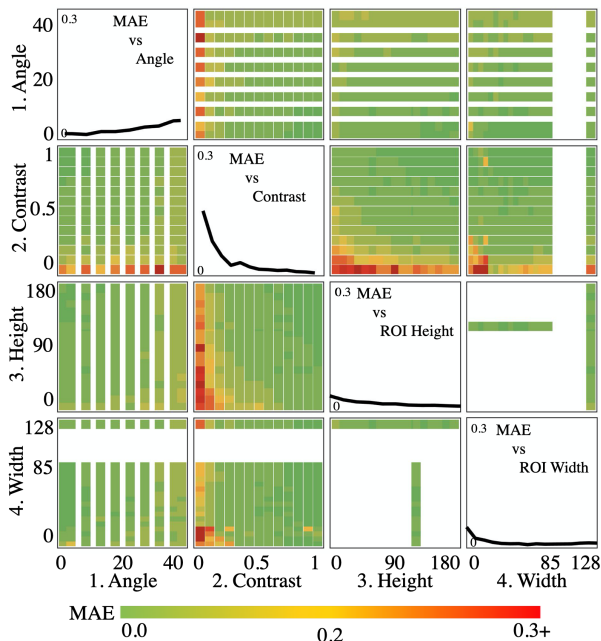

Figure 1. Matrix illustrating the variance introduced to the slanted-edge e-SFR measurement when changing edge angle, edge contrast, ROI height and width. The diagonal of the matrix plots the MAE introduced by each SFR parameter.
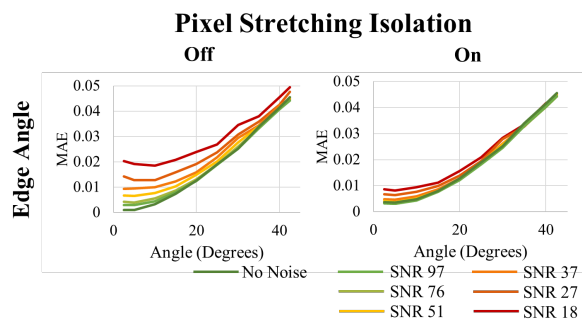

Figure 2. MAE introduced in the slanted-edge SFR measurement with changing edge angle with and without pixel stretching.

4. The NS-SFRs derived with parameter ranges presented in Table 1 are averaged, per radial distance, in the spatial domain. This is achieved by aligning the maxima of their resampled natural scene LSFs and taking a mean at each sample point. The averaged LSF is converted into the frequency domain via the Fourier transform, providing an averaged NS-SFR. These mean NS-SFRs form the six system e-SFR estimates across the frame. The weighted mean of these system e-SFR estimates is calculated, again in the spatial domain, obtaining an overall system e-SFR estimate [15]. This weighted mean is applied to the six radial distance system e-SFR estimates to eliminate bias due to areas of high-density NS-SFRs.

System e-SFR Estimation Results

The methodology presented above was implemented using two image datasets, each taken with a single camera system. The first system consisted of a Nikon D800 DSLR, equipped with a 24mm lens set at f/4. The second was the Apple iPhone7 smartphone camera. The DSLR and smartphone camera datasets included 1800 and 2000 images, respectively. They contained images captured using the same optical focal length and aperture but various shutter speeds and ISO gain settings. The captured scenes varied in content and illumination. They included urban and rural architecture, indoor and outdoor scenes, and various nature scenery, with forests, beaches, and mountains.

NS-SFR data was gathered from each image in the dataset according to the framework presented in [5]. The data from the entire dataset was compiled to estimate the system e-SFR. Using NS-SFR data from many diverse images achieved two traits to allow a more robust measure for the development of this proposed method. Firstly, it improved the chances of obtaining edges from optimal step-edge inputs, reducing the scene component of the NS-SFRs. Secondly, it minimised the potential of missing data in the radial distance segments.

DSLR Camera System

RAW files from this system were converted into 16-bit TIFF files, with sharpening and denoising turned off in the demosaicing process. In most research applications, the TIFF file is adequate for system SFR/MTF measurement since it is considered to incorporate minimum non-linear ISP. In addition to the TIFF files, the green channel of the mosaiced RAW files (sensor images) was used for comparison.

The ISO 12233 slanted-edge method [3] was employed to characterise the system e-SFR. The RAW images of the captured test chart were converted to TIFF files in the same manner as the captured natural scenes dataset. The mean SFR obtained from the target's edges and the standard deviation was calculated for each radial distance, providing the target ISO12233 e-SFR across the frame. The weighted mean of the average SFRs from all six radial distances was calculated to represent the system ISO12233 e-SFR of the entire frame.

The weights used in this instance were 1.00 for the centre, 0.75 for the partway regions and 0.50 for the corners of the frame. They correspond to the default weights in Imatest software employed for SFR analysis [15] but can be adjusted depending on the application. For example, image quality metrics apply a higher weight in the frame's corners than in the centre (higher weights assigned to the poorer SFRs) [16].

The DSLR system e-SFR estimates, obtained from the demosaiced TIFF natural image files and the mosaiced RAW files, were compared to the ISO12233 e-SFR in Figure 3. This figure illustrates the vertical system e-SFR estimate for three radial distance segments: the centre (1/6), partway (3/6) and the corners of the frame (6/6), and also the weighted mean of the entire frame. The left column shows the system e-SFR estimate, the middle the absolute error from the mean ISO12233 e-SFR and the last the radial distance.

In this instance, comparing the system e-SFR estimates to the ISO12233 e-SFR assesses the accuracy of the method. Excluding the high-frequencies, the system e-SFR estimates derived from TIFF and RAW file types stay within or close to the ISO12233 e-SFR standard deviation limits.

The system e-SFR estimated from the TIFF versions of the image files is consistently higher than that of the RAW counterpart, which shows a closer match to the ISO12233 e-SFR. However, the RAW estimates contain a boosted high-frequency SFR, a known image noise attribute [17]. These system e-SFR signatures indicate that denoising is in the TIFF pipeline.

Smartphone Camera System

JPEG files from the smartphone camera were used, meaning that the dataset contains artefacts from compression and a non-linear ISP. Effects from such processes can be observed in the system SFR in Figure 4. This figure illustrates the vertical mean e-SFR and the standard deviation envelope (obtained from the ISO 12233 test chart), along with the corresponding system e-SFR estimate. Figure 4 also illustrates the texture-MTF, calculated using the Imatest spilled-coins test chart [18], typically used to access camera performance with high ISP, for further comparison [8].

Isolated step-edges from ISO 12233 test charts are prone to heavy sharpening, denoising and compression, boosting the system e-SFR, especially at low spatial frequencies. Sharpening is not as effective when extracting step-edges from complex natural scene images due to surrounding scene content and textures, so the low-frequency boost is reduced. Additionally, noise reduction in textured areas will be greater than isolated test chart edges, reducing the estimated system e-SFRs. This non-linear behaviour is reflected in the estimate having a higher association with the texture-MTF in the low frequencies.

These observations indicate that the system e-SFR estimate can potentially be a scene-dependant performance measure, unlike the e-SFR. However, further work is required to establish the full impact of non-linear ISPs on the measure.

Discussion and Conclusions

The proposed methodology for estimating the system e-SFR is accurate and comparable to the standardised ISO 12233 slanted-edge method for linear camera systems. This reliability makes it ideal for several vision systems, including autonomous vehicles and live security systems.

However, the method has caveats that need to be addressed to allow for a live-SFR measure. The current computation time is extensive, taking on average 20 minutes per DSLR 36.3-megapixel image. This computation time is far from a minimum of 24 images per second required for live-SFRs to characterise cameras incorporated in real-time systems. Thus, to work with such systems, the proposed algorithm calls for optimisation. First, lowering frame resolution, reducing it to typical magnitudes for real-time systems, would significantly improve computation times. Incorporating a trained CNN for ROI localisation and validation of step-edges from natural scenes would allow the framework to select the edges most likely to conform to the selected ROI parameters (Table 1). This approach would reduce the time spent processing edges that are subsequently not included in the system e-SFR estimation.
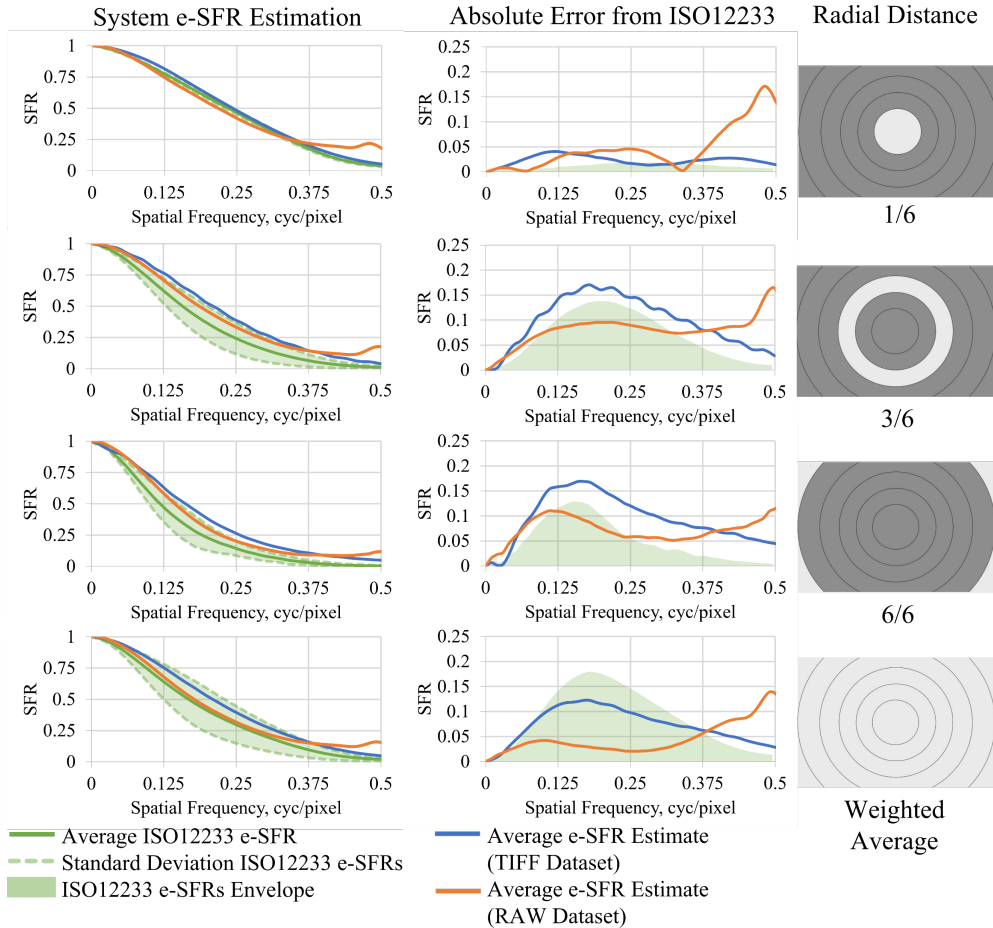
Figure 3. *The vertical system e-SFR estimation for three radial distances out of six and a weighted mean of all six radial distances. The first column contains the estimated e-SFR in relation to the ISO12233 e-SFR. The second column contains the absolute error of these estimated system e-SFRs. The third column contains a visual representation of the radial distance the data belongs.*

System e-SFR results here are produced from compiling 1800-2000 images with very diverse scene content. To produce a live-SFR, the number of images required would have to be reduced. It is worth noting that in autonomous vision systems, such as autonomous vehicles, the camera input signals are unlikely to be as diverse as this study's image datasets. Thus,
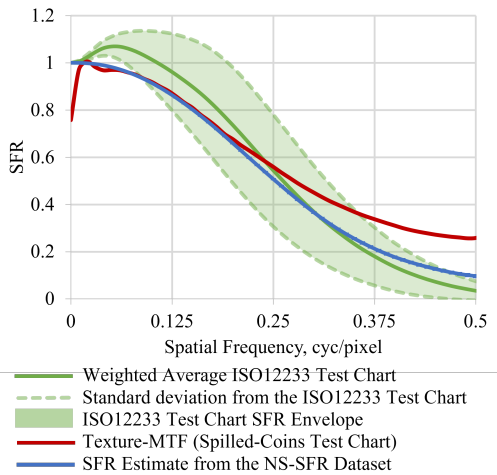

Figure 4. *This graph shows the vertical system e-SFR estimation from the iPhone7 database. This data is compared to the vertical e-SFR envelope measured from an ISO 12233 test chart and the texture-MTF calculated from the spilled-coins test chart.*

further research is currently being carried out to assess the reliability of the method with fewer images, along with more targeted scene contents.

In summary, this paper proposes a method to estimate the camera ISO 12233 e-SFR directly from natural scene SFRs (NS-SFRs). These estimates were taken from the NS-SFR data with tested SFR parameter ranges (edge angle, edge contrast and ROI size) that gave a stable result without limiting the number of extracted natural scenes edges. The camera frame was divided into radial distance segments to reduce variation introduced across the optical image circle, and the highest performance edges were selected per radial distance segment. The selected NS-SFRs were averaged in the spatial n to form system e-SFR estimates across the frame. A weighted mean was then used to produce the system e-SFR estimate.

The resulting system e-SFR gave good approximations of the ISO 12233 e-SFR, especially for systems with less heavy ISP (as expected from an edge-based method). Non-linear system processes produced an estimated system e-SFR that exhibits signs of a scene dependent nature.

Further work is required to make the proposed measure suitable for implementation in autonomous vision systems. This work includes improving the computational time by implementing a CNN for edge localisation, reducing image resolution to a typical number for such systems, the number of images required for a reliable estimate, and finally determining the types of scenes best suited to the application at hand.

Publications

[1] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6, doi: 10.1109/QoMEX.2016.7498955.

[2] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *ICLR*, pp. 1–16, 2019.

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

[4] O. Van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Edge detection techniques for quantifying spatial imaging system performance and image quality," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 1871–1879, 2019, doi: 10.1109/CVPRW.2019.00238.

[5] O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Camera System Performance Derived from Natural Scenes," in *IS&T Electronic Imaging: Image Quality and System Performance XVII*, 2020, pp. 241-1-241–10(10), [Online]. Available: https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-241.

[6] B. Jayachandraiah, S. Jayabharathi, S. Muralikrishnan, and A. S. Kumar, "Development of Advanced Data Quality Evaluation Tools for High Resolution IRS-Satellite Data with IDL / ENVI," *ENVI/IDL User Mtg.*, pp. 1–8, 2005.

[7] L. He, P. Michel, S. Puglia, and D. Williams, "Computer Assisted Image Analysis for Objective Determination of Scanning Resolution for Photographic Collections – An Automated Approach," *Proc. Arch. Conf., IS&T*, no. 1, pp. 43–47, 2013.

[8] P. D. Burns, "Refined Measurement of Digital Image Texture Loss," *Proc. SPIE Vol. 8653, Image Qual. Syst. Perform. X*, vol. 8653, 2013, doi: 10.1117/12.2004432.

[9] R. Branca, S. Triantaphillidou, and P. Burns, "Texture MTF from images of natural scenes," *Soc. Imaging Sci. Technol.*, pp. 113–120, 2017.

[10] E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, J. R. Jarvis, and R. E. Jacobson., "Validation of Modulation Transfer Functions and Noise Power Spectra from Natural Scenes," *J. Imaging Sci. Technol.*, 2019.

[11] M. Bauer, V. Volchkov, M. Hirsch, and B. Sch, "Automatic Estimation of Modulation Transfer Functions," in *IEEE International Conference on Computational Photography (ICCP)*, 2018, pp. 1–40.

[12] D. Williams, "Benchmarking of the ISO 12233 Slanted-edge Spatial Frequency Response Plug-in," in *IS&T's PICS Conference*, 1998, pp. 133–136.

[13] S. Birchfield, "Reverse-Projection Method for Measuring Camera MTF," *Soc. Imaging Sci. Technol.*, pp. 105–112, 2017.

[14] J. K. M. Roland, "A study of slanted-edge MTF stability and repeatability," *SPIE Proc. Vol. 9396*, pp. 1–9, 2015, doi: 10.1117/12.2077755.

[15] Imatest, "Using SFR part 2: Running Imatest SFR." https://www.imatest.com/docs/sfr_instructions2/ (accessed Dec. 04, 2020).

[16] BSI Standards Publication, "BS ISO 20462-3:2012 Psychophysical experimental methods for estimating image quality Part 3: Quality ruler method," *BSI Stand. Publ.*, pp. 1–23, 2012.

[17] E. C. Yeadon, R. A. Jones, and J. T. Kelly, "Confidence Limits for Individual Modulation Transfer Function Measurements based on the Phase Transfer Function," *Photogr. Sci. Eng.*, vol. 14, no. 2, pp. 153–156, 1970.

[18] Imatest, "Spilled-Coins, Dead Leaves, and Random Chart Analysis Scale-invariant test charts (v.2021.1)," *Imatest*, 2021. http://www.imatest.com/docs/random/ (accessed Jul. 05, 2021).