

Analysis of Natural Scene Derived Spatial Frequency Responses for Estimating Camera ISO12233 Slanted-edge Performance

Oliver van Zwaneberg[^], Sophie Triantaphillidou[^], and Alexandra Psarrou
School of Computer Science and Engineering, The University of Westminster, UK
E-mail: o.vanzwaneberg@my.westminster.ac.uk

Robin B. Jenkin[^]
NVIDIA Corporation, USA
School of Computer Science and Engineering, The University of Westminster, UK

Abstract. *The Natural Scene derived Spatial Frequency Response (NS-SFR) framework automatically extracts suitable step-edges from natural pictorial scenes and processes these edges via the edge-based ISO12233 (e-SFR) algorithm. Previously, a novel methodology was presented to estimate the standard e-SFR from NS-SFR data. This paper implements this method using diverse natural scene image datasets from three characterized camera systems. Quantitative analysis was carried out on the system e-SFR estimates to validate accuracy of the method. Both linear and non-linear camera systems were evaluated. To investigate how scene content and dataset size affect system e-SFR estimates, analysis was conducted on entire datasets, as well as subsets of various sizes and scene group types. Results demonstrate that system e-SFR estimates strongly correlate with results from test chart inputs, with accuracy comparable to that of the ISO12233. Further work toward improving and fine-tuning the proposed methodology for practical implementation is discussed. © 2021 Society for Imaging Science and Technology. [DOI: 10.2352/J.ImagingSci.Technol.2021.65.6.060405]*

1. INTRODUCTION

This paper presents an evaluation of a methodology presented in Ref. [1] that automatically extracts camera Natural Scene derived Spatial Frequency Responses (NS-SFRs) from a camera dataset. The NS-SFRs are computed from extracted natural scene step-edges via the ISO12233 e-SFR algorithm [2]. Since the input step edges contain an unknown and varied spatial frequency content, the resulting NS-SFR is a function of the system performance and scene content; separating the two is not straightforward without either being characterized. The proposed method estimates the system e-SFR performance component through statistical evaluation, and extracts the most likely NS-SFR data to match the ISO12233 e-SFR.

[^] IS&T Members.

Received Aug. 7, 2021; accepted for publication Nov. 2, 2021; published online Nov. 18, 2021. Associate Editor: Marius Pedersen.

1062-3701/2021/65(6)/060405/12/\$25.00

This paper provides a quantitative benchmark of the methodology presented in Ref. [1] and elucidates findings from previous studies published [3–5], all by the same authors. The system e-SFR estimation accuracy and precision are reported, and its compatibility with the standard is discussed. Data analysis presented helps to understand the advantages and disadvantages of the approach.

In section 2, a short overview of the NS-SFR framework and the estimated e-SFR from images of natural scenes is presented. Section 3 describes the capture of three diverse image datasets, each from a different camera system. These are used as inputs to estimate the e-SFR of each system. The first two systems are based on a Digital Single Lens Reflex (DSLR) camera with controlled signal processing, so they are considered close to linear. The third system, a smartphone camera, incorporates a highly non-linear image signal processing (ISP). Section 4 presents how the camera datasets are categorized by scene location and divided into sub-sets of various sizes. Results demonstrate accuracy and precision of the system e-SFR estimation with respect to different sized inputs and scene types. The final section addresses the effectiveness of the proposed method for estimating e-SFRs from natural scenes and proposes recommendations for further optimization and refinement of the results.

2. ESTIMATING ISO12233 E-SFR FROM NS-SFR DATA

This section provides a short overview of Ref. [1], i.e., the methodology used to extract NS-SFRs from natural scene captures and the subsequent use of this data to estimate the ISO12233 camera system e-SFR [1].

2.1 Natural Scene Derived e-SFR

NS-SFRs are measured using a series of processes to detect, isolate and verify step-edges from natural scene images.

First, the Canny edge detector [6] is used to locate image edges. Then, the resultant binary edge map is used to crop regions of interest (ROIs) with appropriate

window sizes. A pixel stretching filter was developed to improve the yield of edges isolated from the images [1, 4]. This filter removes unwanted artifacts from either side of the edge of interest, such as noise, scene textures and illumination non-uniformities, shown to confound the resulting e-SFRs [1]. Each ROI then goes through a step-edge verification process [1], analyzing the edge gradient to determine if the edge of interest is of the correct profile. Finally, selected ROIs are processed with the standard slanted-edge algorithm to obtain a series of NS-SFRs. Burns' *sfrmat4* is utilized for this purpose [7].

2.2 Camera System e-SFR Estimation

NS-SFRs for a single camera system are used to estimate the system e-SFR. Estimation is achieved by minimizing variation in the NS-FSR data and averaging. NS-SFRs are first grouped according to six radial annulus segments across the frame before averaged. This process returns six system e-SFR estimates, one per radial annulus. A weighted average returns an estimate for the entire frame.

A threshold is set to obtain the optimal step-edges from the NS-SFRs, i.e., selecting edge inputs that maintain higher modulation across the spatial frequencies than the camera system's rendering capabilities. This threshold is achieved by

measuring the NS-SFR line spread function (LSF) full width at half maximum (FWHM) distribution per radial annulus, then selecting the narrowest LSFs. The top 10th Percentile of each distribution is used as this threshold.

The data is further refined by isolating NS-SFRs measured from edge and ROIs with relevant parameters set to introduce low variation to the system e-SFR [1, 3].

As a result, only a small number of NS-SFRs are selected per radial annulus, but they are stable and produce accurate system e-SFR estimates across the frame when averaged in the spatial domain. The average was performed in the spatial domain to reduce the high frequency bias caused by image noise [8], which can compile when the SFR is averaged in the frequency domain.

The weighted average of the resulting six radial system e-SFRs provides the overall estimate. The weights used here are 1.00 for the central part of the frame, 0.75 for the partway and 0.50 for the frame's corners. These zone weight values are used as default in the Imatest imaging performance evaluation software [9] but can be adapted depending on the application [10, 11].

3. RESULTS OF SYSTEM E-SFR ESTIMATION

3.1 Image Datasets

Three digital camera systems were used to capture a large number of scenes, producing three corresponding image datasets. In this section, the datasets were used in their entirety for estimating system e-SFRs. These datasets are not necessarily examples of applications but are examples of different system pipeline characteristics. The cameras and settings were as follows:

1. DSLR 1 – Nikon D800, 36.3-megapixel sensor (4.87 μm pixel pitch), with a 24 mm lens at $f/4$ (wide-angle).
2. DSLR 2 – Nikon D800, 36.3-megapixel (4.87 μm pixel pitch) sensor, with a 135 mm lens at $f/4$ (telephoto).
3. Smartphone camera – Apple iPhone7, 12-megapixel sensor (1.22 μm pixel pitch), with a 3.99 mm lens at $f/1.8$.

During capture, the shutter speed and ISO gain settings were set to best expose the images (most images were captured using the camera systems' base ISO gain). As a result, each dataset contained a range of noise levels and thus different signal-to-noise ratio (SNR) distributions. Figure 1 plots the distribution of the ISO gain settings, as reported in the EXIF image metadata.

Each of the three datasets comprise of images that were subsequently classified according to three different scene locations: man-made exteriors, indoor scenes, and nature scenes. The AlexNet convolutional neural network (CNN) [12], after being transfer learned [13], was employed for the classification; the process is detailed in section 4. All three datasets have similar scene type distributions, as illustrated in Figure 2.

The camera systems were characterized using the ISO12233 slanted-edge method [2] to obtain lab-based

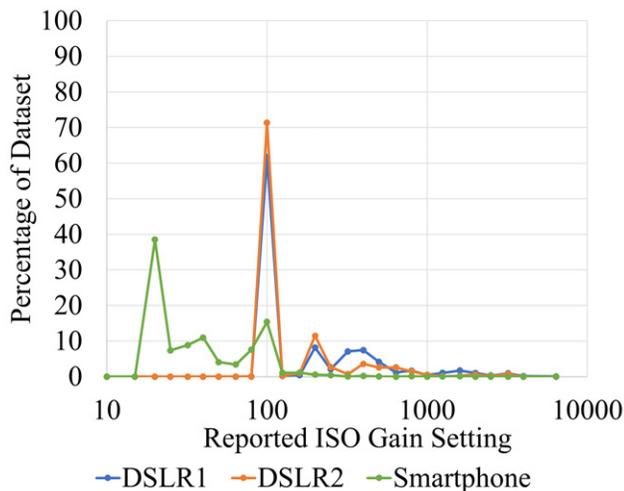


Figure 1. The distribution of the reported ISO gain settings for the images in the three datasets.

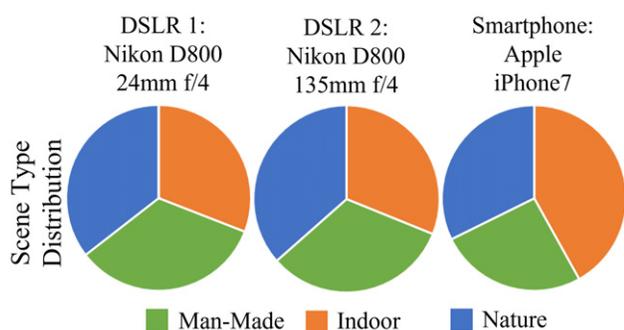


Figure 2. The scene type distributions for the three databases, classified by re-trained AlexNet.

e-SFRs across the camera frame. The frame was divided into six radial annuli in the same fashion as in the e-SFR estimates from natural scenes (section 2.2). ISO12233 e-SFRs of the six segments were also averaged using the same weighted values as the system e-SFR estimation. These ISO12233 e-SFRs are considered the measurement against which estimates are assessed.

For each average ISO12233 e-SFR measure, the standard deviation was calculated to provide a range deemed acceptable when assessing the accuracy of the estimated system e-SFRs, i.e., if the estimation is within the precision range of the equivalent test chart measure.

The DSLR 1 camera dataset contains 1866 images. The wide-angle focal length of this system allows for wide depth of fields, thus reducing the number of out of focus edges. It is worth noting that wide-angle lenses tend to have a wide variation in performance across the imaging circle. This allows the proposed approach to be assessed against significant optical e-SFR variations across the field of view. Two versions of this dataset were produced: the first comprises of uncompressed TIFF image files without sharpening or denoising applied in the demosaicing process; the second comprises of the green channel of the RAW sensor image.

The DSLR 2 camera dataset contains 1009 images. Although their camera body is identical, the two DSLR systems have different optical system performance characteristics due to their lens designs. In contrast to wide-angle lenses, telephoto optics tend to have a smaller performance variation across the frame. Telephoto lenses provide a shallower depth of field, producing many edges out-of-focus, which are not suitable for accurate system e-SFR estimations. This second dataset tests the estimation method's robustness against tighter optical performance tolerances for each radial annulus and its effectiveness in eliminating defocused edges from the calculation. Only uncompressed TIFF files were used, again, with the denoising and sharpening turned off during demosaicing.

The smartphone camera dataset, containing 2008 images, was set to test the suitability of the proposed method for assessing the performance of non-linear pipelines. Smartphone camera images are subject to heavy non-linear ISP and JPEG compression. Both processes are scene dependent and introduce different artifacts to the captured edges.

3.2 Region of Interest Utilization

The highly selective process yields small numbers of optimal natural scene step-edges. After the deselection process to obtain stable NS-SFR data described in section 2.2, it was found that, on average, only 3.41% of the natural scene isolated ROIs were utilized in the system e-SFR estimation. This percentage corresponds to 1.69 ROIs per image. The LSF FWHM threshold is the greatest elimination factor, isolating approximately 10% of the NS-SFRs. The edge parameter thresholds further restrict the NS-SFRs, with the greatest

Table I. The number of test chart and natural scene step-edges isolated per radial annulus segment for each system to measure/estimate the average e-SFR.

Radial annulus	System						
	DSLR 1			DSLR 2		Smartphone	
	ISO12233	RAW	TIFF	ISO12233	TIFF	ISO12233	JPEG
1	14	189	371	15	155	16	110
2	30	328	649	33	324	32	246
3	43	498	937	42	423	42	348
4	41	689	953	55	439	44	479
5	31	258	765	50	294	29	269
6	5	79	236	17	93	5	75

constraint coming from the edge contrast range 0.55–0.65 Michelson contrast [1], reducing the remaining data by 66%.

The number of isolated edges across the frame is not evenly distributed; the center and corners of the field of view have on average 80% fewer isolated ROIs than the partway segments. This uneven distribution is due to the changing segment area and is shown to impact the accuracy of the system e-SFR estimation across the frame. Table I lists the numbers of ROIs used per radial annulus for each system.

3.3 System e-SFR Estimation Results

In this section, estimated system e-SFR results are presented for radial annulus segments 1/6, 3/6 and 6/6, as well as for the weighted average of the frame. These results provide a suitable overview of the performance of the method.

Two camera performance metrics were calculated to analyze the accuracy: the 25% and 50% of Nyquist frequency. Figure 3 contains the delta values between the ISO12233 e-SFR mean value with each of these system e-SFR estimate summary metrics. It is important to note that a single metric cannot fully describe the SFR function. For example, if an estimate passes through the ISO12233 e-SFR at either 25% or 50% of Nyquist frequency, the absolute error at this point will diminish, not describing the error of the entire function.

3.3.1 DSLR “Linear” Camera Systems

Figures 4 and 5 describe the vertical system e-SFRs derived from the DSLR 1 and DSLR 2 datasets, respectively. Horizontal estimates show similar trends and are therefore not included. In both Figures, the first column plots the estimated system e-SFR in relation to the equivalent ISO12233 e-SFR radial annulus mean and ± 1 standard deviation (e-SFR envelope, indicating the “acceptable” error range - see section 3.1). The second column shows the absolute error from the mean ISO12233 e-SFR, with the ISO12233 e-SFR standard deviation envelope again shaded. The third column depicts the radial annulus segments from which the data is measured. Combined with the data presented in Fig. 3, these plots suggest moderate to high accuracy in the system e-SFR estimations.

All natural scene derived system e-SFRs are shown to be approximations of the ISO12233 e-SFR. The RAW

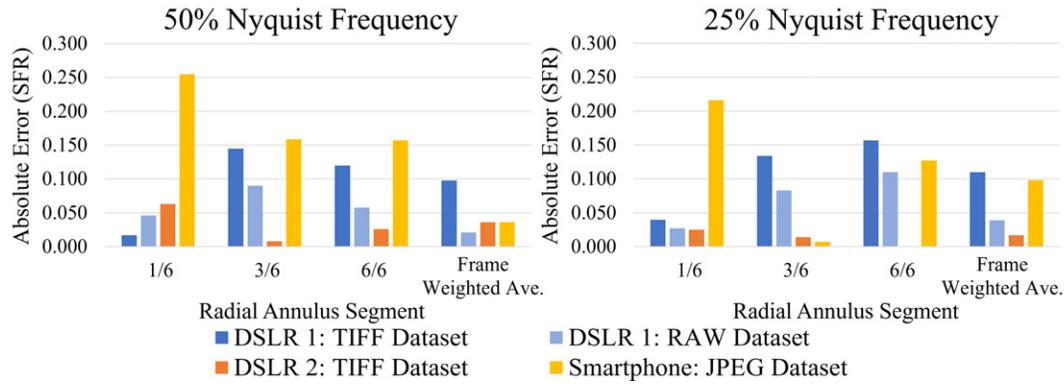


Figure 3. Absolute error at the 50% and 25% system e-SFR estimate Nyquist frequencies in relation to the corresponding ISO12233 e-SFR.

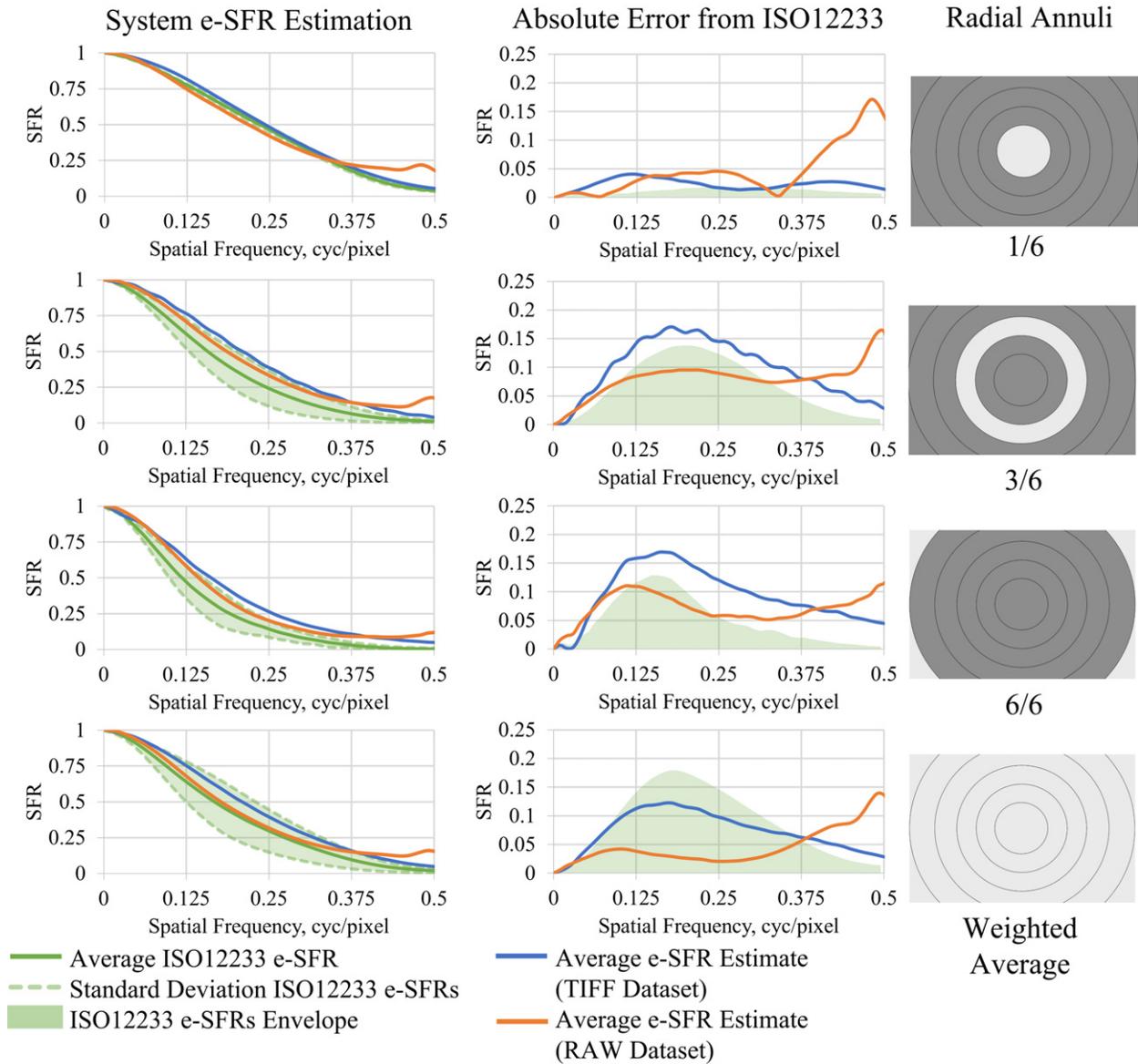


Figure 4. The DSLR 1 vertical system e-SFR estimation for three radial annuli out of six and a weighted mean of the entire field of view. The first column contains the estimated system e-SFR in relation to the ISO12233 e-SFR. The second column contains the absolute error between the estimated system e-SFR from the mean of the ISO12233 e-SFR. The third column contains a visual representation of the radial annulus from which the data belongs.

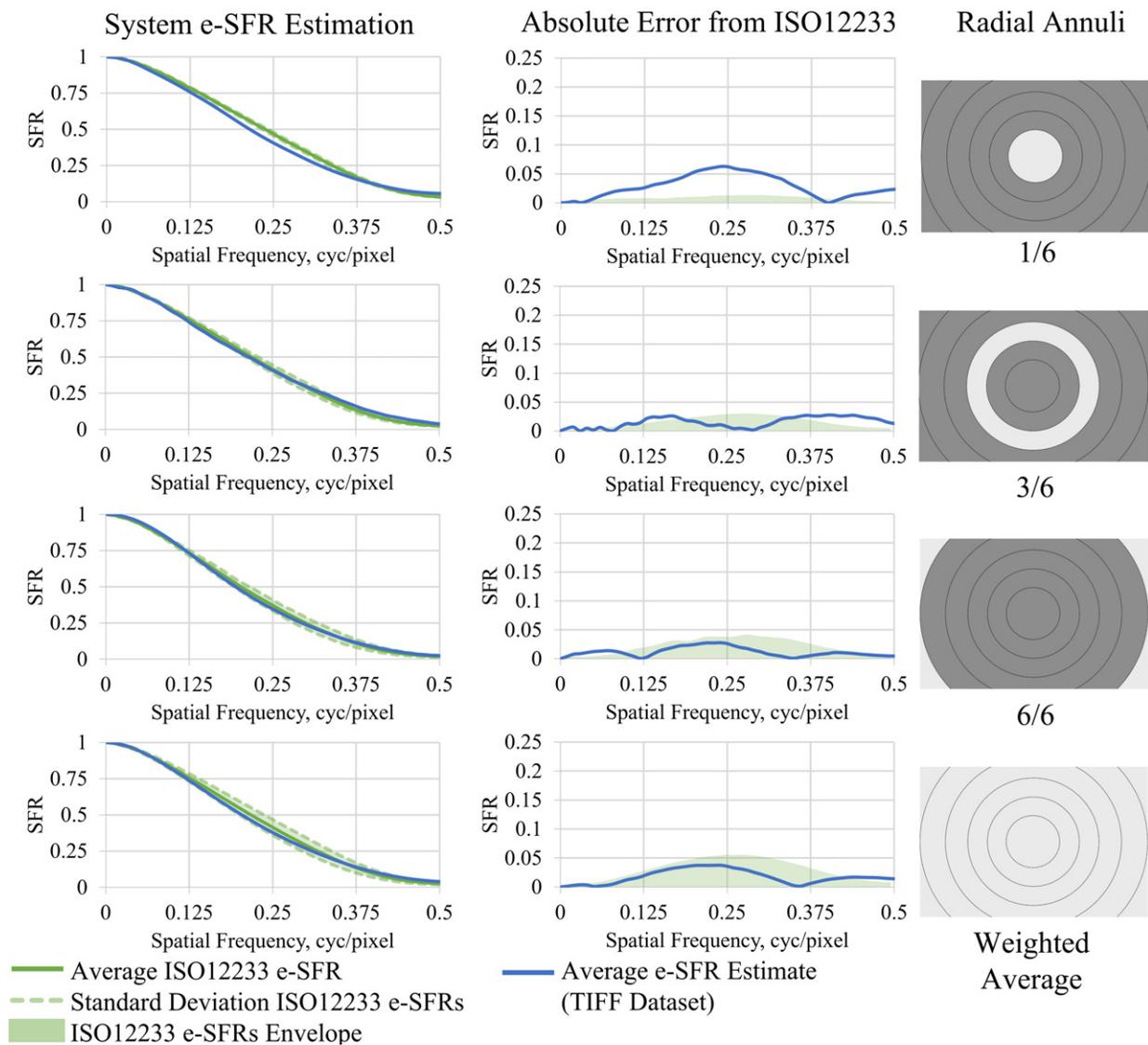


Figure 5. The DSLR 2 vertical system e-SFR estimation for three radial annuli out of six and a weighted mean of the entire field of view. The first column contains the estimated system e-SFR in relation to the ISO12233 e-SFR. The second column contains the absolute error between the estimated system e-SFR from the mean ISO12233 e-SFR. The third column contains a visual representation of the radial annulus from which the data belongs.

DSLR 1 dataset returned a more accurate system e-SFR than the TIFF counterpart. High frequencies were overestimated throughout all RAW estimated system e-SFRs, a known bias associated with high noise levels [8]. DSLR 1 TIFF results do not show this bias, indicating the TIFF pipeline likely incorporates denoising. Also, it is observed that the TIFF results, although follow the same trend as the ISO12233 e-SFRs, are overestimated throughout the spatial frequencies.

The DSLR 2 dataset returned system e-SFRs with higher accuracy than DSLR 1, suggesting that the proposed method is robust against shallow depth of field and is proficient with obtaining telephoto lens performance with tight precision tolerances.

The weighted average estimates use all isolated ROIs across the frame. For both systems, the resultant system

e-SFRs are in accordance with the average ISO12233 e-SFR and within the standard deviation limits, thus providing the most consistent estimate of the system e-SFR from natural scene inputs.

3.3.2 Smartphone “Non-Linear” Camera System

The smartphone system e-SFR estimations for radial annuli 1/6, 3/6 and 6/6, as well as the frame weighted average, are plotted in Figure 6 with the corresponding ISO12233 e-SFRs and standard deviation envelopes. Figure 7 depicts the smartphone weighted average estimate in relation to both the texture-MTF (i.e. dead leaves MTF), measured using the Imatest spilled-coins test chart [14], and the ISO12233 e-SFR. The texture-MTF is designed to simulate an average natural scene signal, resulting in a more faithful camera response for systems subjected to non-linear processing.

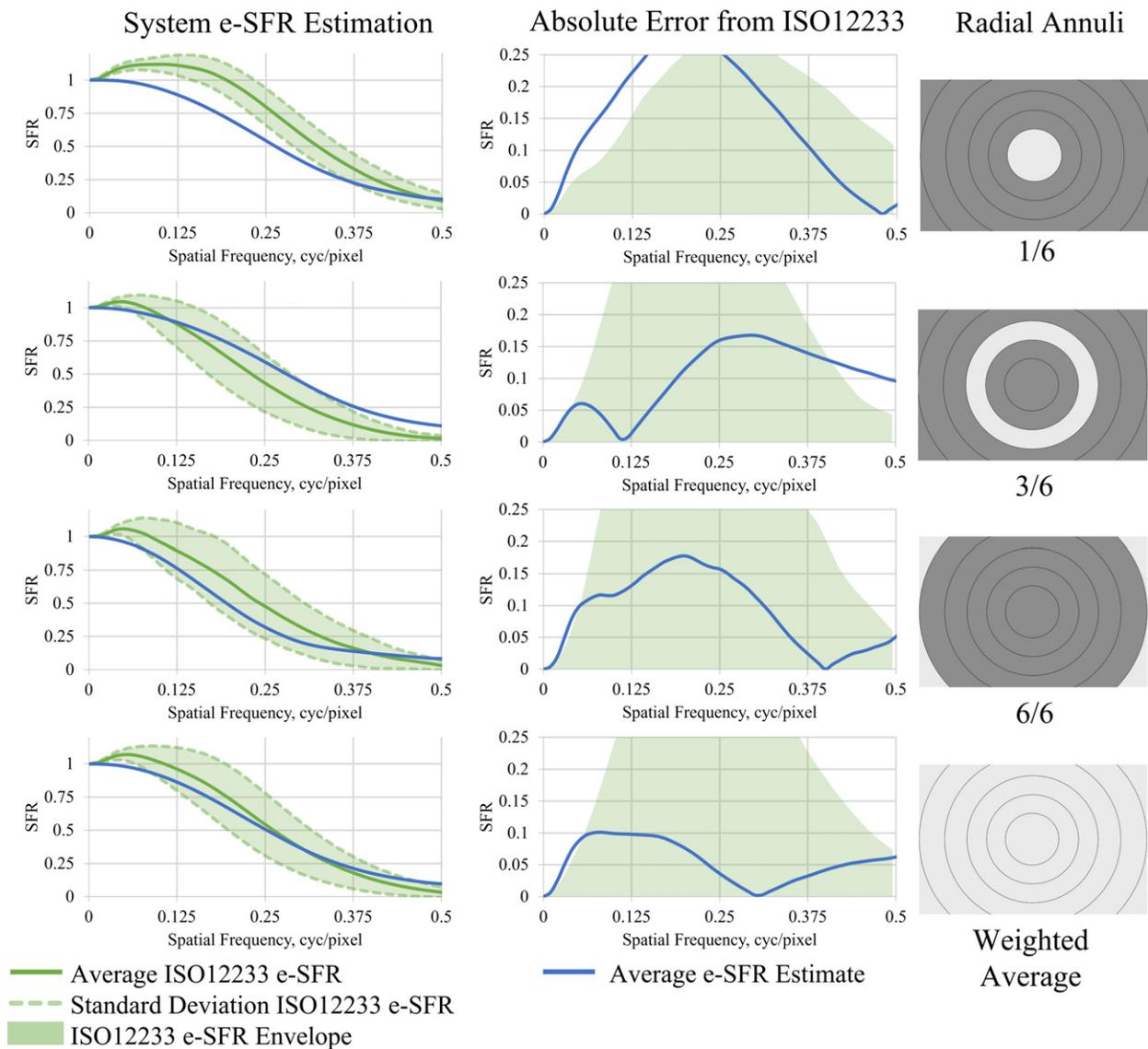


Figure 6. The Smartphone vertical system e-SFR estimation for three radial annuli out of six and a weighted mean of the entire field of view. The first column contains the estimated system e-SFR in relation to the ISO12233 e-SFR. The second column contains the absolute error between the estimated system e-SFR from the mean ISO12233 e-SFR. The third column contains a visual representation of the radial annulus from which the data belongs.

Scene dependency originating from non-linear processing is observed in the results. When isolating step-edges from test charts, adaptive processing (e.g., sharpening, denoising, compression) has an insignificant effect; chart edges are preserved and enhanced, resulting in a low frequency boost in the e-SFR. Processing step-edges in complex natural scenes does not result in a sharpening boost; the inclusion of surrounding scene content and textures means that sharpening is not that effective on natural scene edges. Denoising is also less effective in textured natural scene images compared to isolated test chart edges; thus, image noise adds bias to the estimation.

The smartphone camera dataset was segmented into three scene classes to investigate scene-dependent effects

in the system e-SFR estimates. These were man-made (exterior), indoor and nature scenes, as defined in section 4. Vertical e-SFRs for each dataset are plotted in Figure 8. Horizontal system e-SFR estimations follow the same trends.

Results show that individual class estimated system e-SFRs fall within the ISO12233 e-SFR acceptable limits but do not show the boost in high-frequencies observed in the ISO12233 e-SFR. Man-made scene class has a boosted mid-frequency, indicating stronger sharpening in this class than in the other two. This is an expected trend since nature scenes do not have strong step-edges but busy textures. Indoor scenes may contain step-edges, but generally have lower and non-uniform illumination levels and higher gain settings resulting in image noise, which prompts more

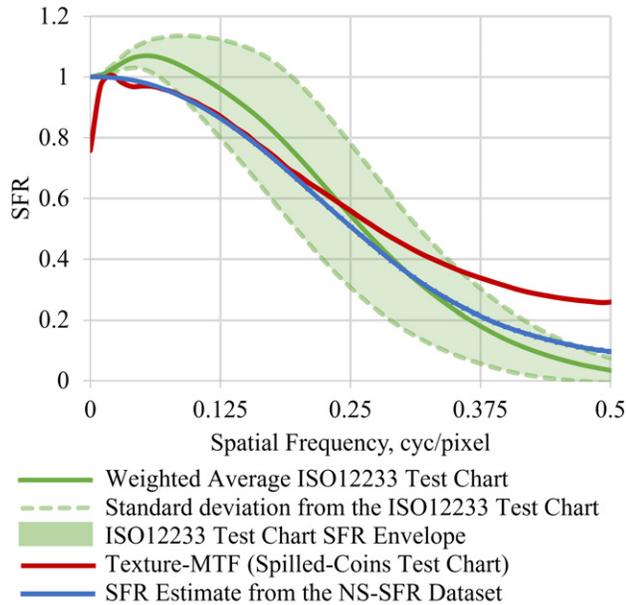


Figure 7. The weighted average vertical system e-SFR estimate from the entire Smartphone dataset. These estimates are compared to the ISO12233 slanted edge method and the Spilled-Coins Texture-MTF.

denoising, resulting in blur and further reduction of the system e-SFR estimate. High-frequency textures and high system noise levels prompt image processing that results in similar estimated e-SFR system responses.

4. VARYING THE NUMBER OF IMAGES IN E-SFR ESTIMATION

Images with a variety of scene contents were used to estimate the system e-SFR. Such diverse inputs are appropriate in certain use cases, for instance, distinction between consumer and professional camera systems. Some situations, such as monitoring real-time camera performance for autonomous vision systems [3], may require input images more specific to the application. This section investigates the effects of restricting scene content as well as the number of images from which the system e-SFRs are estimated.

The DSLR 1 RAW image dataset was divided into smaller subsets and specific scene categories for the purpose. The rationale for using the RAW sensor images rather than TIFF was that the estimates from RAW files, at 25% and 50% Nyquist frequency, have higher accuracy. More importantly, the TIFF data showed signs of non-linear processing, potentially skewing the results.

DSLR 1 performance varies considerably across the field of view. In contrast, DSLR 2 has a greater variation at shallow depth of field. These system characteristics make separating the image number error from the imaging circle location variation or depth of field variation difficult. Furthermore, the effect of a shallow depth of field on the e-SFR estimate becomes greater with fewer images to select the optimal edges. Thus, the DSLR 1 was used in this study, color coding the estimations according to the ISO12233 e-SFR

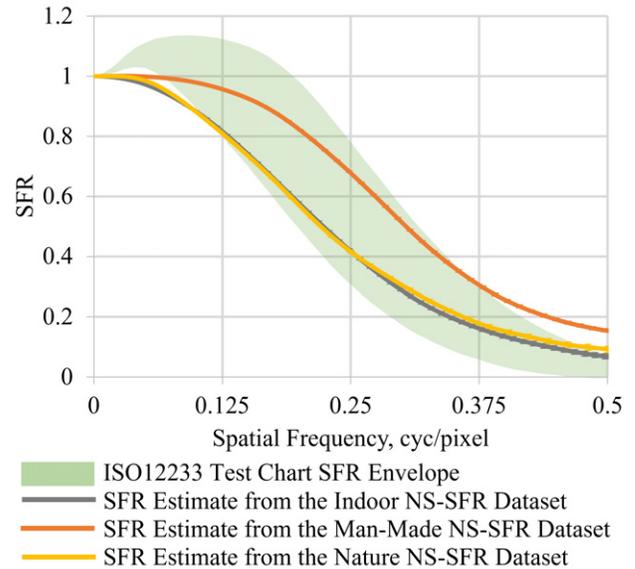


Figure 8. The weighted average vertical system e-SFR estimate from the Smartphone sub-datasets of indoor, man-made and nature scene classes. These estimates are compared to the ISO12233 slanted edge method standard deviation envelope.

standard deviation to depict the acceptable variation from an equivalent test chart measurement.

4.1 Scene Classification

The DSLR 1 images were split into three scene classes:

- (i) Man-made exteriors, urban or rural that include built structures.
- (ii) Indoor scenes that include interiors of shelters, homes, and public buildings.
- (iii) Nature scenes, which include landscapes, plants and animals.

This simple classification allowed the breakdown of scenes in groups containing (i) well-defined step-edges with low noise and low scene texture, optimal for the slanted edge method, (ii) defined edges but with lower contrasts and lower illumination levels, thus containing higher image noise levels, and (iii) less defined step-edges, containing more natural textures, suboptimal for the algorithm. It is to be noted that within each class, there was still a diversity in SNRs, scene objects and illumination.

Scene classification was achieved through transfer learning [13] the AlexNet CNN [12]. The last three layers (fully connected, softmax and classification layers) were rebuilt in the learning process.

An extensive database containing 8574 images was used in the transfer learning process. Half of these images were captured for this purpose. The other half was taken from two open-access databases: SUN [15] and Natural Scene Statistics in Vision Science [16, 17].

Images were split 70:30 for training and validation purposes, respectively. The overall classification accuracy was high, at 96.35%. For a secondary validation, 100

Table II. Seven subdivisions of the scene classed DSLR 1 dataset.

DSLR 1 dataset subdivisions	Number of images per subset (m)	Number of subsets per scene class (n)
1	10	57
2	20	28
3	40	14
4	80	7
5	160	3
6	320	1
7	570 (all available)	1

DSLR 1 images per class were used to calculate the top-1 error rate (the percentage of the first prediction label was incorrect) [18]. The top-1 error rate for the man-made classification was 18%, for the indoor classification 6%, and for the nature classification 2%. The highest source of error was found in scenes containing elements from multiple classes.

4.2 Subdividing the Scene Classes

The images in the three scene classes of the DSLR 1 dataset were divided into n groups containing m randomly selected images; these groups are referred to as subsets. This subdivision process was repeated seven different times, each time doubling m , as listed in Table II. As m increased, the number of subsets, n , decreased.

These subsets were used to evaluate the number and type of scenes necessary for obtaining system e-SFR estimates comparable to the standard. The absolute error at 50% Nyquist frequency was calculated between the individual subset estimates and the equivalent ISO12233 e-SFR for this purpose.

4.3 Results

The absolute error of the individual subset estimations was analyzed with respect to subset size and scene class. Figure 9 provides a sample of this data analysis, plotting the absolute error of the subdivided DSLR 1 RAW vertical system e-SFR estimates plotted for radial annuli 1/6, 3/6 and 6/6, and the weighted average of the entire frame. The data points are color-coded to indicate whether the estimates are within, above, or below the ISO12233 e-SFR standard deviation limits. The mean absolute error (MAE) was calculated by averaging the subset error data points per subdivision. The MAE for each subdivision is plotted to show accuracy trends as the number of images processed through the proposed method increases.

Figure 10 plots the missing data and the standard deviation for each of the seven subdivisions against the radial annulus segment. Missing data is a percentage of subsets that did not provide an estimate for a given radial annulus segment. The standard deviation plots show the spread of

the estimates for each subdivision, providing a precision measurement. As subdivisions 6 and 7 only contain one subset, thus their standard deviations cannot be calculated.

Three traits affect the accuracy of the estimation, which change across the field of view with the number of images used in the estimation:

- (i) Quantity of edges available for the e-SFR estimate,
- (ii) the scene dependency and suitability of the natural scene step-edges [5],
- (iii) and the natural performance variation of the system.

The center of the frame, 1/6, has the least precise estimates, seen across all three scene classes, with high standard deviations. This is due to the presence of few isolated ROIs, as previously shown, resulting in missing data. For subdivision 1, this missing data reaches up to 45% for the man-made and indoor classes and 70% for the nature class, i.e., it is difficult to estimate a center radial segment system e-SFR using a small subset. When this segment contains suitable step-edges, a larger subset improves the accuracy and precision of the system e-SFR estimation, as seen in man-made and indoor scene classes. Center nature scene class e-SFR estimates, with higher missing data and suboptimal step-edges, have no improvement in either accuracy or precision.

This same observation is made for the corner frame segment, which has a maximum of 80% missing data due to few isolated ROIs and step-edges with lower suitability for all three scene classes. Thus, there is no improvement in precision with larger subsets. This is the lowest performance frame segment for a system due to high optical distortion and artifacts, such as chromatic aberrations, which are typically avoided within test chart based measurements. In contrast to the other frame segments, a larger subset decreases the accuracy, seen clearly in the nature class. More images provide more data to select optimal natural scene step edges, and the NS-SFR LSF FWHM distribution will impact the 10th percentile threshold. This is important in the corners of the frame, as artifacts such as chromatic aberrations can distort the measurement, increasing the distribution. Larger subsets have a higher likelihood to contain data with more artifacts, negatively impacting the estimate by increasing the distribution of the LSF FWHMs. In addition, location of corner edges will impact the estimation. The corners of the frame, particularly for a wide-angle lens, will have the greatest natural variation in the e-SFR. Small subsets may not be evenly distributed across each corner, resulting in subsampling the range optical e-SFR. Incorporating more images (thus, more edges) will provide more even distribution, resulting in the average of the range of e-SFRs, thus the increase in MAE for this radial annulus. This is a challenging position in the field of view for system e-SFR estimation with missing data and problematic optical characteristics.

The partway region, 3/6, yields a high number of ROIs. As a result, fewer subsets contain missing data for the region. The standard deviation generally decreases with more

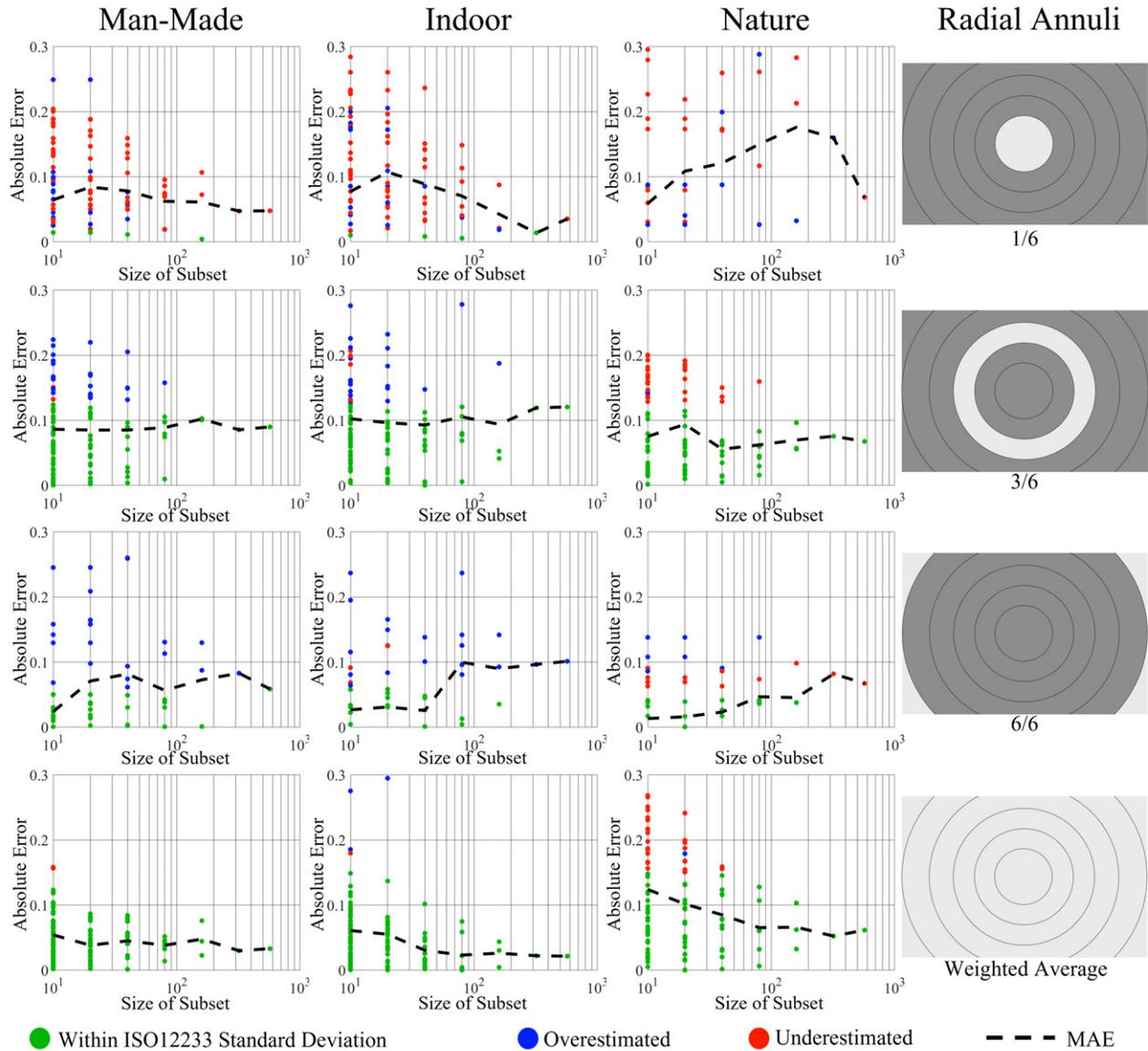


Figure 9. Plots depicting the precision and accuracy of the 50% Nyquist system e-SFR estimation, at radial annuli 1/6, 3/6, 6/6 and the weighted average of the frame, using various sized subsets. This data is presented for three scene classes, man-made (exterior), indoor and nature.

images per subset, seen clearly in the man-made scene class. Although the standard deviation decreases, the MAE values stayed constant across the seven subdivisions for all three scene classes. In other words, when there are enough ROIs isolated, the average accuracy stays constant, but precision improves with more images. Moreover, the MAE stays within the ISO12233 e-SFR standard deviation, providing system e-SFR estimates comparable to the standard for all three scene classes.

The weighted mean system e-SFR estimation shows optimal conditions, with rich data from every subset. As a result, the standard deviation and MAE decrease with more images per subset, i.e., precision and accuracy improve. The accuracy improvement is minor in scenes with suitable step-edges. However, in a scene with greater scene textures, the precision and accuracy improvements are far greater.

Providing a generalized conclusion of how many images are required for an e-SFR estimate is impractical, as it is application dependent; instead, the number of edges should be determined using the following criteria:

- (i) The scene content - Studying the scene classes, as expected, the system e-SFR is estimated with higher accuracy using small datasets containing images with well-defined step-edges. When using suboptimal scene inputs, it is beneficial to use large datasets and/or the weighted average estimate to assess system performance.
- (ii) The e-SFR requirements - In the event the application requires an e-SFR across the entire field of view. In that case, more images are required for accurate e-SFR estimates in the center and corner frame segments. In

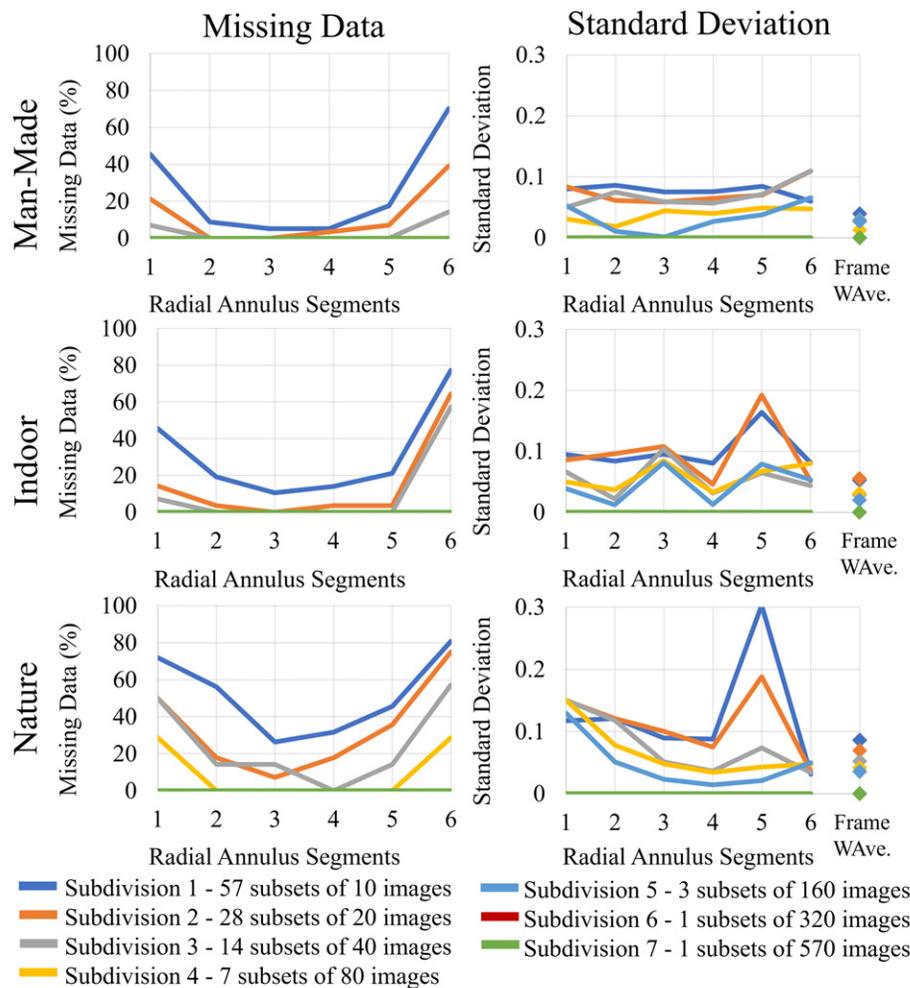


Figure 10. Plots illustrating the missing data and standard deviation across the seven DSLR 1 subdivisions against the radial annulus segments per scene class.

contrast, if a single overall estimate is sufficient, fewer images provide an accurate e-SFR estimation across the tested scene classes.

- (iii) The rate of change/variance of content between each image - In an application where the input images have variety or are constantly changing, few images are required as the edges would be distributed across the entirety of the field of view. On the other hand, if the signal feed is static, increasing the number of images would not be beneficial.
- (iv) The camera system - This study has used a high pixel resolution DSLR camera system. A lower pixel sensor and different optical characteristics would result in different step-edge yields and thus would require different numbers of images to achieve a robust e-SFR estimation.

Keeping these criteria in mind, assuming the ROI yields are similar to the DSLR 1, the following are examples of generalized practical applications. An automotive vehicle in an urban setting would have numerous man-made structures. While driving the movement of edges would provide edges distributed across the field of view. If an overall

frame average is required, 10 images would be adequate to stay within the precision of the test chart e-SFR. This would provide three estimates every second, assuming a 30 frames per second video feed.

For the same example system but now requiring the e-SFR variation across the field of view, 80 images would balance the precision and accuracy in the center and corner frame segments. This would provide the e-SFR estimates every 2.67 seconds.

Another example would be a CCTV system. Such systems would have a static, or slow moving, video feed of man-made or indoor structures. Due to the inability to obtain evenly distributed edges across the field of view, a weighted average e-SFR of the frame would be most suited. Alternatively, the exact field of view location e-SFRs may be more appropriate rather than using azimuthal averages per annulus.

5. DISCUSSION

In this paper, an accurate system e-SFR estimate is shown to be achievable from NS-SFR data.

It shows that several conditions should be considered when implementing the proposed method in Ref. [1] to optimize the system e-SFR estimate.

Demosaiced image files are adequate for estimating the system e-SFR, provided the ISP is linear or close to linear. Otherwise, the RAW image can be used to ensure linearity. If the application allows, an extensive image dataset was shown to increase the estimate's accuracy. When the optical e-SFR variation across the frame is not required, a weighted frame average provides higher accuracy and precision than each of the six individual radial annuli. Alternatively, the number of segments can be decreased, for instance, using three segments rather than six.

Most modern cameras are non-linear, with the ISP treated as a black-box and the RAW image not easily accessible. Further research is required to determine the scene-dependent nature of the NS-SFRs and subsequent system e-SFRs estimates for highly non-linear systems since they vary with scene contents. A study that breaks down localized scene content combined with the localized NS-SFRs would further this understanding. In addition, researching the NS-SFRs from different edge contrast ranges and using wider LSF FWHM distribution thresholds (e.g., 10th to 20th percentiles) would be beneficial. Since the e-SFR is edge-based, a scene and process-dependent measure would be more sensitive to changes in non-linear sharpening than other processes.

In addition, further work is required to tailor the method to a specific application. The algorithm implemented here is not optimized with respect to the processing time. Currently, e-SFRs from natural scenes are derived through a two-stage process, deriving the NS-SFR and then deselecting the unwanted edges to achieve the most appropriate data. As discussed in Ref. [1], combining these two steps, and selecting the appropriate edges from the start will allow the proposed method to be used effectively and leading to the estimated system e-SFR being derived in real-time, embarking many application opportunities, for example, in autonomous vision systems [3].

The proposed method depends on the natural scene edge input; many blurred edges would result in underestimations due to changing the shape of the LSF FWHM distribution, resulting in the selection of wider LSFs in the 10th percentile threshold. An alternative approach would be to take a statistical analysis of the NS-SFR extrema. Further work should look into developing the proposed method to estimate the system e-SFR through NS-SFR statistics.

6. CONCLUSIONS

The proposed methodology in Ref. [1] for estimating camera system e-SFRs from image datasets of natural scenes is shown in this paper to provide a close approximation of the ISO12233 e-SFR. The weighted average estimate stays within one standard deviation of the ISO12233 measured result for linear or close-to-linear camera systems.

Results from a smartphone system incorporating highly non-linear ISP indicate that the e-SFR estimate is scene and

processing-dependent. Further work is required to establish correlations between specific scene contents (especially scene edge types, texture and noise) and e-SFR estimates from natural scenes. Such a study would provide insight into scene-and-process dependent camera ISP behavior and on whether e-SFRs estimated from natural scenes would be more suitable camera performance measures for non-linear camera systems than the lab-based ISO 12233 e-SFR.

The distribution of the derived edge ROIs across the frame showed consistency for all tested systems. The ROI utilization, i.e., the NS-SFR ROIs used in the system e-SFR estimate, is shown to be low, at 3.41% on average. In a dataset of natural scenes, most of the sequesterable data came from the partway field of view segments. The center and corners of the field of view have considerably lower yields. As a result, this may lead to missing data when implementing the proposed approach in a 40 image dataset or smaller, or for images without many step-edges, such as scenes of nature.

Using the weighted average across the frame segments mitigates some of the low ROI yield error since it combines all suitable ROIs from the dataset. The result is a more stable and accurate system e-SFR estimate than the individual radial annuli. Thus, when measurement across the imaging circle is not required, it is recommended to use the weighted average, adjusting the weight values to suit the application [10, 11].

The study included the DSLR 1 dataset being classed into scene classes and subdivided into subsets (i.e., smaller input image datasets). Estimating the system e-SFR from subsets of different scene types and sizes provided essential observations. The larger image subset increased accuracy and precision for radial annulus segments found to have a low number of isolated step-edges, such as in the center segment. In radial annuli with poorer, fewer and less distributed step-edges, such as the frame corners, precision may increase, but accuracy can decrease with larger subsets due to optical distortion artifacts negatively impacting the step-edges. Partway radial annuli were shown to contain a sufficient number of isolated step-edges; these areas maintain the same average accuracy across the subset sizes and for all scene classes, but the precision increases with larger datasets. The weighted average was also shown to give the most stable result for small subsets, staying within the ISO12233 standard deviation limits, although accuracy and precision improve using larger subsets.

As expected, the data also showed that scene types with more suitable step-edges do derive better system e-SFRs. These scenes, e.g., man-made and indoor, allow estimates that maintain accuracy within one standard deviation of the ISO12233 using small datasets across the radial annulus segments. For suboptimal scenes, e.g., nature, the weighted average estimate would provide the best results increasing the precision.

To conclude, results from implementing the proposed method using three different camera systems are overall encouraging. With further work to improve code usability and processing time optimization, the proposed method can be implemented in several computer vision and imaging

science applications, allowing accurate estimates of the ISO 12233 e-SFR directly from natural scenes.

MATLAB CODE

The latest version of the MATLAB code and the three image databases used to estimate the system e-SFR can be found at: <https://github.com/OlivervZ11/NS-SFR>.

Note, some of the images and all number plates have been removed on data protection grounds.

ACKNOWLEDGMENT

We wish to show our appreciation to Dr Jae Young Park for funding this research project.

REFERENCES

- ¹ O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Estimation of ISO12233 edge spatial frequency response from natural scene derived step-edge data," *J. Imaging Sci. Technol.* **65**, 060402-1–060402-16 (2021).
- ² British Standard Institute "BS ISO 12233:2017 Photography - Electronic still picture imaging - Resolution and spatial frequency responses," (BSI Standards Publication, 2017), pp. 1–62.
- ³ O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Natural scene derived camera edge spatial frequency response for autonomous vision systems," *Proc. IS&T London Imaging Meeting* (IS&T, Springfield, VA, 2021), pp. 88–92.
- ⁴ O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Camera system performance derived from natural scenes" *Proc. IS&T Electronic Imaging: Image Quality and System Performance XVII* (IS&T, Springfield, VA, 2020), pp. 241-1–241-10.
- ⁵ O. Van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Edge detection techniques for quantifying spatial imaging system performance and image quality," *IEEE: New Trends in Image Restoration and Enhancement (NTIRE) Workshop, in Conjunction with Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2019), pp. 1871–1879.
- ⁶ J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698 (1986).
- ⁷ P. D. Burns, "sfrmat4 source code," (Burns Digital Imaging, 2020), (accessed Oct. 04, 2020).
- ⁸ E. C. Yeadon, R. A. Jones, and J. T. Kelly, "Confidence limits for individual modulation transfer function measurements based on the phase transfer function," *Photographic Sci. Eng.* **14**, 153–156 (1970).
- ⁹ Imatest, "Using eSFR ISO Part 2 (v.2021.1)," 2021, (accessed Dec. 07, 2020).
- ¹⁰ British Standard Institute, "BS ISO 20462-3:2012 Psychophysical experimental methods for estimating image quality Part 3: Quality ruler method," (British Standard Institute, 2012), pp. 1–23.
- ¹¹ B. Keelan, *Handbook of Image Quality: Characterization and Prediction*, 1st ed. (CRC Press, New York, 2002).
- ¹² A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inform. Process. Syst.* **1**, 1097–1105 (2012).
- ¹³ L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications*, edited by E. Soria, J. Martin, R. Magdalena, M. Martinez, and A. Serrano (IGI Global, Hershey, 2009), pp. 242–264.
- ¹⁴ Imatest, "Spilled-coins, dead leaves, and random chart analysis scale-invariant test charts (v.2021.1)," *Imatest*, 2021, (accessed Jul. 05, 2021).
- ¹⁵ J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "SUN Database: Large-scale scene recognition from Abbey to Zoo," *IEEE Conf. Computer Vision Pattern Recognit.* (IEEE, Piscataway, NJ, 2010).
- ¹⁶ W. S. Geisler and J. S. Perry, "Statistics for optimal point prediction in natural images," *J. Vis.* **11**, 1–17 (2011).
- ¹⁷ University of Texas at Austin, "Natural Scene Statistics in Vision Science Image Database," (2020).
- ¹⁸ D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *Int'l. Conf. on Learning Representations (ICLR)* (arXiv, New York, 2019), pp. 1–16.