

CameraVDP: Perceptual Display Assessment with Uncertainty Estimation via Camera and Visual Difference Prediction

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 1077

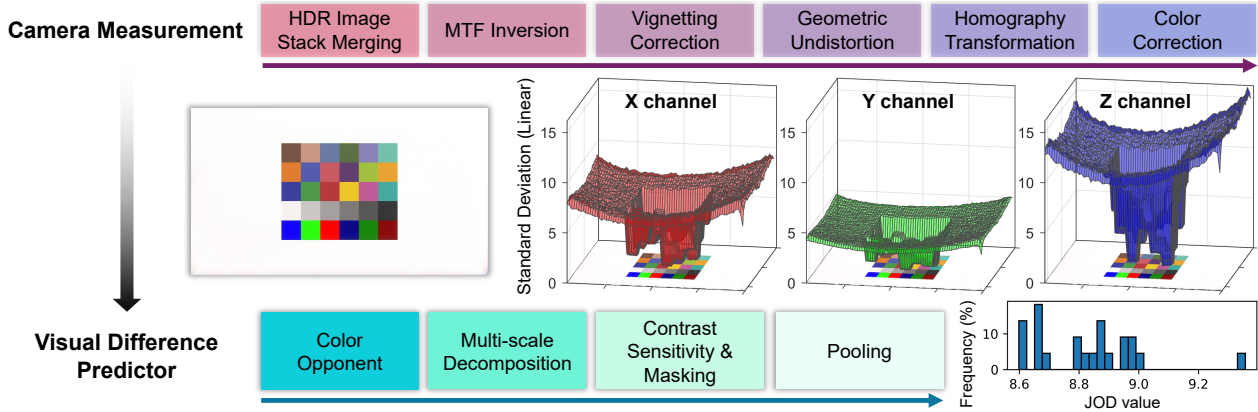


Fig. 1. CameraVDP pipeline. The camera first captures distortion-free, color-corrected measurements along with their uncertainty covariance (standard deviations in XYZ space shown). Test and reference measurements are then input to the visual difference predictor to compute a quality score distribution. Values and uncertainties are jointly propagated at each step. (JOD: just-objectionable-differences, see [Mantiuk et al. 2021, 2024])

Accurate measurement of images produced by electronic displays is critical for the evaluation of both traditional and computational displays. Traditional display measurement methods based on sparse radiometric sampling and fitting a model are inadequate for capturing spatially varying display artifacts, as they fail to capture high-frequency and pixel-level distortions. While cameras offer sufficient spatial resolution, they introduce optical, sampling, and photometric distortions. Furthermore, the physical measurement must be combined with a model of a visual system to assess whether the distortions are going to be visible. To enable perceptual assessment of displays, we propose a combination of a camera-based reconstruction pipeline with a visual difference predictor, which account for both the inaccuracy of camera measurements and visual difference prediction. The reconstruction pipeline combines HDR image stacking, MTF inversion, vignetting correction, geometric undistortion, homography transformation, and color correction, enabling cameras to function as precise display measurement instruments. By incorporating a Visual Difference Predictor (VDP), our system models the visibility of various stimuli under different viewing conditions for the human visual system. We validate the proposed CameraVDP framework through three applications: defective pixel detection, color fringing awareness, and display non-uniformity evaluation. Our uncertainty analysis framework enables the estimation of the theoretical upper bound for defect pixel detection performance and provides confidence intervals for VDP quality scores.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGGRAPH 2025, Conference Location,
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/XXXXXXX.XXXXXXX>

CCS Concepts: • **Computing methodologies** → **Computational photography; Perception.**

Additional Key Words and Phrases: visual difference predictor, computational displays, spatial color measurement

ACM Reference Format:

Anonymous Author(s). 2025. CameraVDP: Perceptual Display Assessment with Uncertainty Estimation via Camera and Visual Difference Prediction. In . ACM, New York, NY, USA, 10 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

Accurate display evaluation is critical for the design and maintenance of display systems, ensuring reliable visual content reproduction across diverse usage scenarios while avoiding visible artifacts. This entails three key requirements: precise display measurement, perceptual modeling incorporating human visual characteristics, and uncertainty estimation in both measurement and modeling.

Display measurement must accurately capture patterns of high or low spatial frequencies to support both pixel-level evaluation (e.g., defective pixels) and full-screen assessment (e.g., uniformity). The ideal solution is to use high-resolution 2D photometers or colorimeters; however, such devices are not affordable for small research laboratories, and their processing stack is proprietary. High-resolution photographic (prosumer) cameras that capture RAW images offer a practical alternative, but they face several challenges: (1) with only three color filters, they cannot reconstruct the input optical power spectrum; (2) camera lenses introduce geometric distortions; (3) their dynamic range is limited; (4) vignetting introduces non-uniformity in captured images; (5) and high spatial frequencies are attenuated or lost due to optical aberrations. To overcome these limitations, we propose an open-source measurement pipeline for

mirrorless cameras, customized for the measurement of displays. It takes advantage of regular display pixel layout and a limited set of subpixel primaries. The pipeline integrates HDR image stack merging, MTF inversion, vignetting correction, geometric undistortion, homography transformation, and color correction.

A display defect or artifact matters only when it is visible to the human eye. The visibility of such artifacts can be predicted by visual difference predictors [Daly 1992; Mantiuk et al. 2024]. Those, however, require perfectly aligned and physically calibrated test and reference images. The goal of our camera correction pipeline is to provide such images for visual difference predictors.

A measurement instrument needs to provide an estimate of the measurement error it makes in order to ensure it is accurate enough for a given task. We include this capability in our method by analytically modeling the propagation of sensor noise (uncertainty) through all the stages of camera image correction. Furthermore, we train multiple-versions of VDP (ColorVideoVDP, [Mantiuk et al. 2024]) on the XR-DAVID dataset to estimate its prediction error via Monte Carlo simulation. This allows us to use the CameraVDP to robustly detect defective pixels, estimate the visibility of color fringing artifacts and display non-uniformity.

In summary, our contributions are as follows:

- An Open Source pipeline¹ for correcting distortions introduced by mirrorless cameras, which takes advantage of display-specific priors. Such a pipeline can be combined with a VDP to estimate the visibility of the distortions.
- A forward uncertainty propagation framework, incorporating analytical normal distribution-based estimation for camera measurement uncertainty and Monte Carlo simulation for VDP uncertainty.
- We demonstrate CameraVDP’s accuracy and the importance of uncertainty estimation through three applications: defect pixel detection, color fringing, and uniformity assessment.

2 RELATED WORK

Geometric and radiometric camera calibration. Geometric and radiometric camera calibration integrate models of optics, sensors, and image processing to bridge the gap between raw sensor measurements and physically accurate image representations. HDR imaging [Debevec and Malik 2023; Hanji et al. 2020] reconstructs relative scene radiance from multi-exposure sequences, preserving details in both shadows and highlights. Color correction [Finlayson et al. 2015] maps RGB values captured with camera primaries into trichromatic color values for a given color matching functions (e.g., CIE XYZ 1931). Geometric distortion correction [Tang et al. 2017; Weng et al. 1992] relies on parametric models to rectify lens distortions. The camera spatial frequency response (SFR/MTF) [Burns et al. 2022] (ISO 12233 standard), typically estimated using the slanted-edge method, quantifies the spatial resolution of imaging systems and serves as a key metric for optical performance evaluation. This work integrates these calibration techniques into a unified pipeline, enabling commercial cameras to accurately capture high-fidelity content for HDR displays, supporting downstream tasks such as

display uniformity and color fringing assessment. We rely on an explainable analytical model rather than a differential network [Tseng et al. 2021], as the former is guaranteed to generalize to any content and is more appropriate for accurate camera measurements.

Modeling visible differences. Due to the high cost of subjective human (expert) evaluation, objective computational models of human visual perception are essential. Accurate models typically incorporate prior knowledge of the human visual system (HVS) [Schade Sr 1956] and psychophysical data, enabling better generalization to unseen scenarios. The contrast sensitivity functions (CSFs) characterize the HVS sensitivity—defined as the inverse of contrast detection threshold—to variations in color [Ashraf et al. 2024], luminance [Mustonen et al. 1993], area [Rovamo et al. 1993], temporal frequency [Cai et al. 2024; Watson et al. 1986], and spatial frequency [Barten 2003]. Based on CSF and contrast masking modeling, visual difference predictors (VDPs) such as DCTune [Watson 1993], VDP [Daly 1992], ColorVideoVDP [Mantiuk et al. 2024], and HDR-VDP-3 [Mantiuk et al. 2023] have been developed to address more complex image and video content. Existing VDPs require perfectly aligned test and reference images, both calibrated in physical units. Our camera correction pipeline provides such alignment and calibration, making camera-capture images a suitable input to VDPs.

Uncertainty quantification. Uncertainty is typically categorized into two types: aleatoric (stochastic) and epistemic (systematic) [Der Kiureghian and Ditlevsen 2009]. Its quantification generally falls into two approaches: forward propagation and inverse assessment. Forward propagation estimates overall system uncertainty based on input variability, using methods such as Monte Carlo simulations [Kroese et al. 2013] and surrogate models [Ranftl et al. 2021]. Inverse assessment corrects model bias or calibrates parameters using observed data, exemplified by modular [Kennedy and O’Hagan 2001] and fully Bayesian [Bayarri et al. 2009] approaches. In computational imaging, common approaches to uncertainty quantification include Markov Chain Monte Carlo (MCMC) sampling [Bardsley 2012; Broderick et al. 2020], Bayesian hypothesis testing [Repetti et al. 2019], variational inference [Arras et al. 2019; Blei et al. 2017; Ekmekci and Cetin 2021; Rezende and Mohamed 2015; Sun and Bouman 2021], and Bayesian neural networks [Ongie et al. 2020; Xue et al. 2019].

However, most of these methods target deep learning-based reconstruction tasks. While Hagemann et al. [2022] addresses bias and uncertainty in camera calibration via a resampling-based estimator, a unified framework for uncertainty propagation through HDR merging, geometric undistortion, homography transformation, and color calibration remains absent. Moreover, uncertainty modeling is entirely unexplored in perception modeling.

3 METHODOLOGY

Our approach combines camera-based measurements with a visual difference predictor (VDP). The goal of the camera measurements is to achieve sufficient accuracy by correcting the inherent distortions of standard cameras. The complete workflow, shown in Figure 2, comprises six key steps. Based on these measurements, the VDP incorporates a model of human vision to convert physical differences into perceived visual differences. As with any physical measurement,

¹The software will be released before the publication of this work.

including those obtained using camera sensors, measurement uncertainty is inevitable. Additionally, the VDP introduces prediction uncertainty due to its dependence on the training data. A combined estimation of these uncertainties is essential to assess the reliability of the predicted values.

3.1 Camera correction

Our camera measurement correction pipeline is designed for cameras equipped with pixel-shift (sensor-shift) feature, which increases resolution and reduces noise and demosaicing artifacts. For that reason, we do not model demosaicing in our pipeline. We also assume that the measured display has three primaries (i.e., RGB subpixels).

Uncertainty estimation is based on three assumptions: the noise is independent across (1) exposures, (2) individual pixels, and (3) RGGB camera subpixels.

Our derivation involves extensive notation, with full definitions provided in Table S1 (supplementary materials).

HDR Image Stack Merging and Noise Model. The dynamic range of HDR displays often exceeds that of camera sensors, necessitating the capture of multiple exposures. Merging exposures can also reduce noise (via averaging). The RAW (digital) sensor values \mathbf{I} for a pixel $\mathbf{p} \in \mathbb{R}^2$ can be modeled as [Aguerrebere et al. 2014]:

$$\mathbf{I}_c(\mathbf{p}) \sim k_c \{ \text{Pois}((\Psi_c(\mathbf{p}) + d)t)g + \mathcal{N}(0, \sigma_{\text{read},c}^2)g + \mathcal{N}(0, \sigma_{\text{adc},c}^2) \}, \quad (1)$$

where $c \in \{r, g, b\}$ is the color channel index, t is the exposure time, $g = \text{ISO}/100$ is sensor's gain, d is the dark current noise, and Ψ is the scene radiance. k_c , $\sigma_{\text{read},c}^2$ and $\sigma_{\text{adc},c}^2$ are the noise parameters, which we estimate in Section 4.1. We measured the dark noise d for our test camera by capturing I with the lens blocked ($\Psi \approx 0$) at room temperature. As the dark noise remained below 0.1% within our exposure range, it was omitted from further analysis.

Before merging a stack of RAW exposures $i = 1 \dots N$, we need to compensate for differences in exposure time and gain. The relative radiance $\mathbf{X}_c(\mathbf{p})$ is defined as $\mathbf{X}_c(\mathbf{p}) = \frac{\mathbf{I}_c(\mathbf{p})}{t_i g}$. We can formulate the reconstruction of the scene radiance by incorporating the noise model from Eq. (1) and solving for the maximum likelihood estimation (MLE) [Aguerrebere et al. 2014]. However, solving for MLE is impractical for large sets of high-resolution images. As noted by Hanji et al. [2020], photon noise dominates in modern cameras, especially at long exposure times used for measurements. Therefore, we can assume a simplified noise model $\mathbf{X}_c(\mathbf{p}) \sim k_c \text{Pois}(\Psi_c(\mathbf{p})t)/t$ and the corresponding log-likelihood function:

$$\ell(\Psi_c(\mathbf{p})) = \sum_{i=1}^N \left(\frac{x_{i,c} t_i \log(\phi_c(\mathbf{p}) t_i)}{k_c} - \phi_c(\mathbf{p}) t_i - \log \left(\left(\frac{x_{i,c} t_i}{k_c} \right)! \right) \right). \quad (2)$$

The above likelihood has a closed-form estimator [Hanji et al. 2020]:

$$\bar{\Psi}_c(\mathbf{p}) = \frac{\sum_{i=1}^N k_c x_{i,c}(\mathbf{p}) t_i}{\sum_{i=1}^N t_i}, \quad (3)$$

where $x_{i,c}$ denotes a single measurement ($i = 1 \dots N$) of the random variable \mathbf{X}_c . $\ell(\phi_c(\mathbf{p}))$ was introduced to simplify the notation.

Given the sufficient photon count received by the camera during exposure under typical display luminance, the Poisson distribution can be approximated by a normal distribution. Then, the relative

radiance $\mathbf{X}_c(\mathbf{p})$ follows a normal distribution:

$$\mathbf{X}_c(\mathbf{p}) \sim \mathcal{N} \left(k_c \Psi_c(\mathbf{p}), \frac{k_c^2 \Psi_c(\mathbf{p})}{t} + \frac{k_c^2 \sigma_{\text{read},c}^2}{t^2} + \frac{k_c^2 \sigma_{\text{adc},c}^2}{t^2 g^2} \right). \quad (4)$$

The estimated distribution of radiance $\hat{\Psi}$ from exposure stack is:

$$\hat{\Psi}_c(\mathbf{p}) \sim \mathcal{N} \left(\frac{\sum_{i=1}^N k_c x_{i,c}(\mathbf{p}) t_i}{\sum_{i=1}^N t_i}, \frac{\sum_{i=1}^N [k_c^2 x_{i,c}(\mathbf{p}) t_i + k_c^2 \sigma_{\text{read},c}^2 + k_c^2 \sigma_{\text{adc},c}^2 / g_i^2]}{(\sum_{i=1}^N t_i)^2} \right). \quad (5)$$

MTF Inversion. Due to lens aberrations, imperfect focus and lens glare, imaging systems cannot accurately reproduce high spatial frequency details, resulting in blur. The Modulation Transfer Function (MTF), also known as the Spatial Frequency Response (SFR), characterizes an imaging system's ability to resolve fine spatial details. Classical measurement methods include edge-based (e-SFR) [Kerr 2024] and sinewave-based analysis (s-SFR) [Loebich et al. 2007].

The scene radiance $\Psi_c(\mathbf{p})$ affected by blur and glare can be modeled as the convolution of the clean scene radiance $\mathbf{L}_c^M(\mathbf{p})$ and the camera point spread function (PSF) $P(\mathbf{p})$, plus noise $\eta(\mathbf{p})$:

$$\Psi_c(\mathbf{p}) = (\mathbf{L}_c^M * P)(\mathbf{p}) + \eta(\mathbf{p}), \quad (6)$$

where $*$ denotes convolution. The MTF $M(\omega)$ is the modulus of the Fourier transform of the PSF $P(\mathbf{p})$ [Burns et al. 2022], $\omega \in \mathbb{R}^2$ is spatial frequency in cycles-per-pixel. Taking the Fourier transform \mathcal{F} on both sides of Eq. (6) yields:

$$\mathcal{F}(\Psi_c) = \mathcal{F}(\mathbf{L}_c^M)M(\omega) + \eta'(\omega), \quad (7)$$

Considering the noise, we use Wiener deconvolution to obtain the deglared and deblurred estimate of \mathbf{L}_c^M :

$$\hat{\mathbf{L}}_c^M(\mathbf{p}) = \mathcal{F}^{-1} \left(\mathcal{F}(\hat{\Psi}_c) G_c(\omega) \right) (\mathbf{p}), \quad (8)$$

where the Wiener filter is ($*$ is the conjugate operator):

$$G_c(\omega) = \frac{M^*(\omega) S_{\hat{\Psi}_c}(\omega)}{|M(\omega)|^2 S_{\hat{\Psi}_c}(\omega) + N_{\hat{\Psi}_c}(\omega)}. \quad (9)$$

$S_{\hat{\Psi}_c}(\omega)$ and $N_{\hat{\Psi}_c}(\omega)$ represent the power spectral densities (PSD) of the signal and noise, see supplementary materials for details.

The MTF was assumed to be isotropic. We measured $M(\omega)$ using the slanted-edge method (Figure 2, column 2), and fitted using:

$$M'(\omega) = a_1 \exp \left(- \left(\frac{\omega - b_1}{c_1} \right)^2 \right) + a_2 \exp \left(- \left(\frac{\omega - b_2}{c_2} \right)^2 \right), \quad (10)$$

where $a_1, a_2, b_1, b_2, c_1, c_2$ are fitting parameters. To suppress noise amplification, $M(\omega) = \max(M'(\omega), 0.5)$, as shown in Figure 4.

Since the Fourier transform is a linear operation, $\hat{\mathbf{L}}_c^M(\mathbf{p})$ follows a normal distribution. The mean of $\hat{\mathbf{L}}_c^M(\mathbf{p})$ is given by Eq. (8):

$$\mu_{\hat{\mathbf{L}}_c^M}(\mathbf{p}) = \mathcal{F}^{-1} \left(\mathcal{F}(\mu_{\Psi_c}) G_c(\omega) \right) (\mathbf{p}), \quad (11)$$

Under the assumption of spatial white noise, the variance is:

$$\sigma_{\hat{\mathbf{L}}_c^M}^2(\mathbf{p}) \approx \sigma_{\Psi_c}^2(\mathbf{p}) \iint |G(\omega)|^2 d\omega. \quad (12)$$

The derivation can be found in Section S1 of the supplementary.

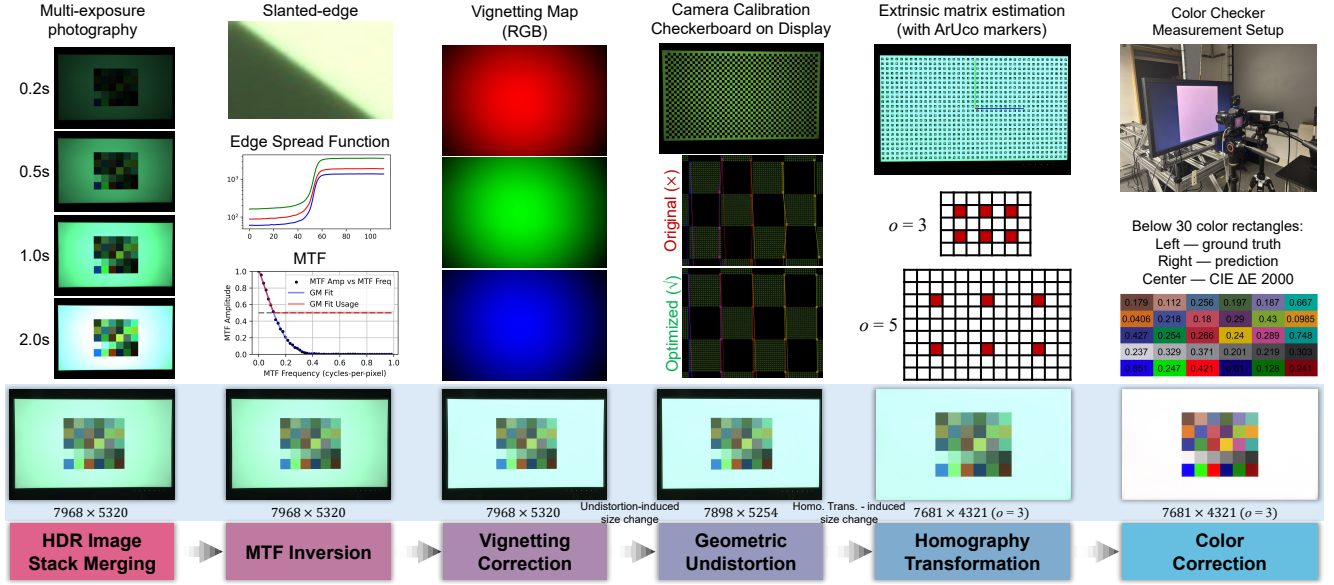


Fig. 2. The complete camera measurement pipeline is shown from left to right, with each column representing a key step and the bottom row illustrating transformation examples. Column 1: HDR acquisition via merging multi-exposure images (4 exposure levels). Column 2: MTF inversion to correct lens aberrations. Column 3: Vignetting correction, exemplified by the vignetting map of the Sony $\alpha 7R$ III with FE 1.8/35 lens. Column 4: Intrinsic and distortion parameter estimation using full-screen checkerboard patterns for geometric undistortion (crop empty edges caused by undistortion). Column 5: Homography estimation of extrinsic parameters using full-screen OpenCV ArUco markers. The display pixel oversampling factor o denotes the side length (in image pixels) of the square formed by the centers of four original display pixels (red squares) in the oversampled image after homography transformation. Examples with $o = 3, 5$ are shown, where red squares indicate display pixel centers. Column 6: Color correction by mapping camera RGB to measured XYZ.

Vignetting Correction. Vignetting refers to the reduction in image brightness toward the edges compared to the center. It can be categorized into mechanical, optical, natural and pixel vignetting, and is associated with factors such as partial occlusion at the lens periphery and the cosine fourth-power falloff [Asada et al. 1996].

Although many reference-free vignetting correction methods [Kang and Weiss 2000; Zheng et al. 2008] exist, flat-field correction remains the most widely used approach. It involves capturing the image I of a uniformly illuminated and color-consistent light source I_{flat} :

$$I_c(\mathbf{p}) = V_c(\mathbf{p})I_{\text{flat},c} + \epsilon_c(\mathbf{p}), \quad (13)$$

where $V_c(\mathbf{p}) \in (0, 1]$ is the vignetting function and $\epsilon_c(\mathbf{p})$ is noise. To compensate for the impact of dust in the optical system, the vignetting function is estimated without smoothing:

$$\hat{V}_c(\mathbf{p}) = \hat{I}_c^M(\mathbf{p}) / \max(\hat{I}_c^M(\mathbf{p})). \quad (14)$$

The vignetting correction result is:

$$\hat{I}_c^V(\mathbf{p}) \sim \mathcal{N}\left(\frac{\mu(\hat{I}_c^M(\mathbf{p}))}{\hat{V}_c(\mathbf{p})}, \frac{\sigma^2(\hat{I}_c^M(\mathbf{p}))}{\hat{V}_c^2(\mathbf{p})}\right). \quad (15)$$

Geometric Undistortion and Homography Transformation.

We aim to find a mapping $m: \mathbf{s} \rightarrow \mathbf{p}$, from screen pixel coordinates \mathbf{s} to camera pixel coordinates \mathbf{p} , such that:

$$\hat{I}_c^D(\mathbf{s}) = \sum_{\delta \in \Omega} \hat{I}_c^V(m(\mathbf{s}) + \delta) R(\delta). \quad (16)$$

where $R(\delta)$ is the resampling kernel and Ω is the neighborhood of a pixel. m is composed of homography $h: \mathbf{s} \rightarrow \mathbf{g}$ and geometric undistortion $u: \mathbf{g} \rightarrow \mathbf{p}$, where \mathbf{g} is the undistorted coordinate.

Homography transformation requires displaying multiple OpenCV ArUco markers [Garrido-Jurado et al. 2014] on screen (see Figure 2, column 5) to obtain the homography matrix $\mathbf{H} \in \mathbb{R}^{3 \times 3}$. Then the screen coordinate can be mapped to an undistorted coordinate with:

$$\mathbf{g}^h = \mathbf{H}\mathbf{s}^h, \quad (17)$$

where $\mathbf{g}^h, \mathbf{s}^h$ are the homogeneous coordinates of \mathbf{g} and \mathbf{s} .

To support applications requiring subpixel color structures (e.g., chromatic aberration), display pixels are supersampled with a factor of o ($o \geq 3$), as shown in the fifth column of Figure 2.

For geometric undistortion, we adopt the Brown-Conrady distortion model, comprising radial and tangential distortions:

$$u(\mathbf{g}) = \mathbf{g} \cdot \left(1 + k_1 \|\mathbf{g}\|^2 + k_2 \|\mathbf{g}\|^4 + k_3 \|\mathbf{g}\|^6 \right) + \left[\frac{2p_1 \mathbf{g}_x \mathbf{g}_y + p_2 (\|\mathbf{g}\|^2 + 2\mathbf{g}_x^2)}{p_1 (\|\mathbf{g}\|^2 + 2\mathbf{g}_y^2) + 2p_2 \mathbf{g}_x \mathbf{g}_y} \right], \quad (18)$$

where k_1, k_2, k_3 are the radial distortion parameters, p_1, p_2 are the tangential distortion parameters, $\|\mathbf{g}\|^2 = \mathbf{g}_x^2 + \mathbf{g}_y^2$.

Using OpenCV for camera calibration on the test display showing a checkerboard pattern, we obtained the intrinsic camera matrix and distortion parameters. Note that the origin of \mathbf{s} is at the display center, with the x -axis pointing right and the y -axis pointing upward, using pixel pitch as the unit.

Additionally, we found that when multiple camera pixels capture a single display pixel (with oversampling), OpenCV fails to accurately detect the corner points. To address this, we first identify all display pixel coordinates via local maxima. Then, the coordinates of the detected checkerboard corners are updated to the average of the coordinates of the two white display pixels (not in the same

checkerboard block) closest to the original detected corner position. (see Figure 2, column 4). When calibrating for the Sony FE 2.8/90 lens on the Eizo display, our improved method reduced the reprojection error from 0.315 to 0.018, achieving an 18-fold improvement compared to the OpenCV calibration (with subpixel refinement).

When estimating uncertainty, the resampling operation introduces local averaging that compromises the independence of neighboring pixels. To simplify subsequent analysis, we conservatively estimate the variance by resampling it in the same manner as the mean (Eq. (16)). Although this may slightly overestimate variance in some regions, it preserves the pixel independence assumption, facilitating Monte Carlo sampling in downstream visual difference predictors.

Color Correction. Cameras cannot capture colors identical to those perceived by the human eye due to differences in spectral sensitivity between the eye and the camera sensor, making color correction essential [Finlayson et al. 2015]. The radiance \hat{L}_c^D recorded by the camera for a single display pixel is:

$$\hat{L}_c^D = \sum_k P_k \int_{\lambda} E_k(\lambda) C_c(\lambda) d\lambda, \quad (19)$$

where λ is the wavelength, $E_k(\lambda)$ is the spectral power distribution (SPD) of subpixel $k \in \{R, G, B\}$ at unit intensity, P_k is the linearized subpixel value, and $C_c(\lambda)$ denotes the spectral sensitivity of camera channel $c \in \{r, g, b\}$. Similarly, the trichromatic CIE XYZ value (related to cone responses) can be expressed as:

$$Y = \sum_k P_k \int_{\lambda} E_k(\lambda) \tilde{S}(\lambda) d\lambda, \quad (20)$$

where $Y = [X, Y, Z]^T$ and $\tilde{S}(\lambda) = [\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)]^T$ are the CIE 1931 standard observer color matching functions (CMFs).

For a display with three color subpixels, there exists a unique transformation matrix $M \in \mathbb{R}^{3 \times 3}$ such that:

$$\hat{Y}(s) = M \hat{L}^D(s). \quad (21)$$

To estimate uncertainty, we need to model covariance as the color channels are no longer independent. For a three-primary (RGB) display:

$$\Sigma_{\hat{Y}}(s) = M \Sigma_{\hat{L}^D}(s) M^T, \quad (22)$$

where Σ is the covariance matrix. See the supplementary material Section S2 for displays with more than three primaries.

3.2 Visual difference predictor

The visual difference predictor (VDP) is a full-reference metric operating on images represented in physical units (luminance, radiance, trichromatic values), which can estimate perceptual differences between test and reference images. We adopt the backbone architecture of ColorVideoVDP [Mantiuk et al. 2024] as our VDP (Figure 3). The inputs are color-corrected camera measurements of the test and reference images (in XYZ space), which are first transformed into the DKL color space (color opponent stage). Each channel (achromatic, RG, YV) is then decomposed into spatial frequency-selective bands using a Laplacian pyramid. The contrast encoded in the bands is used to account for spatio-chromatic contrast sensitivity and contrast masking. Finally, the perceived contrast differences are pooled

into a JOD quality score and a corresponding heatmap. For further details, please refer to the original paper [Mantiuk et al. 2024].

The visual difference predictor involves complex computations, making analytical propagation of uncertainty unfeasible. Additionally, both aleatoric and epistemic uncertainties need to be considered during propagation, corresponding to the image measurement uncertainties from the input test and reference (as discussed in the previous section), and the uncertainty in the parameters of the ColorVideoVDP itself. Therefore, we directly employ Monte Carlo sampling to estimate the uncertainties.

Means and uncertainties of test and reference images derived from the camera measurement pipeline output (\hat{X}) and used to sample 100 test images. To estimate VDP uncertainty, we randomly partition the XR-DAVID dataset [Mantiuk et al. 2024] into 21 train-validation splits and train the model for 6 epochs on each, producing 21 distinct parameter sets. Each configuration is then evaluated on 100 randomly sampled test-reference image pairs, yielding 2100 JOD estimates. The distribution of these JOD values is then treated as the outcome of uncertainty propagation through VDP.

4 APPLICATIONS AND EXPERIMENTS

This section first describes the implementation of camera correction, then uses Monte Carlo simulation to validate the theoretical uncertainty derivation, and finally introduces three applications: defective pixel detection, color fringing, and uniformity perception.

Measurements were conducted on an Eizo ColorEdge CS2740 display (3840×2160) using a Sony $\alpha 7R$ III (ILCE-7RM3) camera equipped with Sony FE 1.8/35 and Sony FE 2.8/90 Macro G OSS lenses. To reduce Bayer demosaicing artifacts, pixel shift multi-shot mode (4-shot) was employed. All analyses were performed on RAW images (.ARQ and .ARW) processed with rawpy. Measurements were acquired at the camera’s full resolution (7968×5320).

4.1 Camera Correction Implementation

Camera noise model parameters. The uncertainty analysis begins with Eq. (1), where we need to estimate k_c , $\sigma_{\text{read},c}^2$, and $\sigma_{\text{adc},c}^2$. From Eq. (1), we have:

$$\sigma_{I_c}^2(\mathbf{p}) = \mu_{I_c}(\mathbf{p}) g k_c + \sigma_{\text{read},c}^2 g^2 k_c^2 + \sigma_{\text{adc},c}^2 k_c^2. \quad (23)$$

Parameter estimation has two stages: (1) capture images of uniform field of varying luminance and fit the linear relationship between $\sigma_{I_c}^2(\mathbf{p})$ and $\mu_{I_c}(\mathbf{p})$ to estimate k_c ; (2) with the camera lens covered ($\mu_{I_c}(\mathbf{p}) \rightarrow 0$), vary g (ISO/100) and fit a quadratic function between $\sigma_{I_c}^2(\mathbf{p})$ and g to estimate $\sigma_{\text{read},c}^2$ and $\sigma_{\text{adc},c}^2$. Results are in Figure 5.

Vignetting. Without a uniform reference (e.g., an integrating sphere), we employed the Eizo display with uniformity correction as a flat-field source. Images were captured using a defocused lens at close range, utilizing only the display’s central region. HDR stacks were used to reduce $\epsilon_c(\mathbf{p})$, and averaging across multiple distances removed subpixel artifacts. $\hat{V}_c(\mathbf{p})$ are shown in Figure 2, column 3.

Color Correction Matrix. To find color correction matrix M , we measured trichromatic values of 30 colors (X-Rite 24-color ColorChecker and full/half-intensity RGB, Figure 1) with JETI Specbos 1211 spectroradiometer in the center of the display. The color

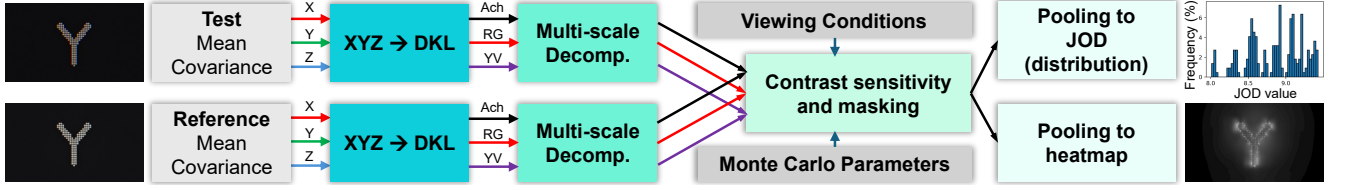


Fig. 3. Visual difference predictor with uncertainty analysis. Input camera measurements (mean and covariance) are transformed to DKL space and used to generate a sample of 100 test images. The VDP pipeline evaluates the difference 21 times, each time using a different set of calibration parameters, to produce the distribution of JOD values and local differences encoded in a heatmap.

Table 1. Results of display defective pixel detection and ablation study of camera measurement pipeline. The bottom row shows our complete pipeline (no color correction). Best PR_{auc} under each condition is highlighted. (HDR: HDR Image Stack Merging. MTF: MTF Inversion. VC: Vignetting Correction. GU: Geometric Undistortion. HT: Homography Transformation.)

Contrast c_{dp}	1			0.2		
Size a_d	4	2	1	4	2	1
HDR-HT	0.967	0.946	0.963	0.987	0.875	0.195
HDR-MTF-HT	0.964	0.945	0.978	0.988	0.885	0.252
HDR-MTF-VC-HT	0.962	0.944	0.976	0.988	0.889	0.213
HDR-MTF-VC-GU-HT	1.000	1.000	0.996	1.000	0.908	0.282

patches were displayed and measured sequentially. Representative color correction results (Sony FE 1.8/35 lens + Eizo display) are shown in Figure 2, column 6. The method achieves high accuracy, with mean color difference CIE $\Delta E_{2000} = 0.308$ between the predicted and ground-truth XYZ values.

4.2 Monte Carlo validation of uncertainty propagation

To validate our analytical uncertainty model from Section 3, we compare it with Monte Carlo simulations. Given the requirement for full-resolution data, this validation is computationally intensive. Using 1000 samples (with mean and variance of HDR-merged captures), we compare theoretical and simulated results in Figure 6, confirming good accuracy of our uncertainty model.

4.3 Detecting defective pixels on a display

Prolonged use of displays may cause subpixel degradation, such as OLED burn-in. Factory defects mechanical damage can also cause defective pixels. CameraVDP precisely captures per-pixel luminance distributions and enables accurate defective pixel detection.

Test patterns. Test patterns simulate a display with defective pixels. Each pattern consists of a uniform white image (pixel value [255, 255, 255]) displayed on a 3840×2160 screen, containing 100 darker square patches simulating defective pixels. All patches within a stimulus share the same edge length $a_d \in \{1, 2, 4\}$ (display pixels) and Weber contrast $c_d \in \{0.2, 1\}$ (corresponding to sRGB-encoded pixel values of 231 and 0) relative to the background. Varying a_d, c_d yields 6 unique test patterns. Experiments were conducted using the Eizo display, the Sony $\alpha 7R$ III camera with the FE 1.8/35 lens.

Detection method. Given parameters a_d, c_d , display size $w_d \times h_d$ (display pixels), and the oversampling factor o used in homography transformation, the measurement resolution is $((o - 1)w_d + 1) \times$

$((o - 1)h_d + 1)$ (image pixels). The mean map M_p is computed by averaging the RGB channels of the output from Eq. (16) (no color correction), then convolving with a box filter of size $(o - 1)a_d - 1$. Similarly, a background mean map M_b is obtained using a box filter size of $10(o - 1) + 1$. A region is labeled defective if $M_p < M_b D_{\text{thr}}$, where detection threshold $D_{\text{thr}} \in [0, 1]$. Redundant detections are removed using non-maximum suppression (NMS) [Canny 1986].

Evaluation metrics. The objective is to identify defective pixels with minimal number of false positives and negatives. The evaluation metric is the area under the precision-recall curve (PR_{auc}), computed across varying detection thresholds D_{thr} . A detection is considered correct if the predicted center lies within a one-display-pixel radius of the ground truth center.

Results. Experimental results in the bottom row of Table 1 show that for $a_d \geq 2$, our method achieves near-perfect $\text{PR}_{\text{auc}} (\approx 1)$. Under extreme conditions ($a_d = 1, c_d = 0.2$), PR_{auc} decreases due to noise (more false positives) as discussed below. Thus, we recommend placing the camera closer to the display to enlarge low-contrast defective pixels in practical use.

Ablation studies. The contribution of each component of our camera correction pipeline to the accuracy of detection is reported in Table 1. The baseline HDR-HT method showed the poorest performance. Incorporating the MTF inversion module (Wiener deconvolution) in HDR-MTF-HT significantly improved results for small defect sizes, confirming the MTF module’s effectiveness and highlighting glare and blur as key challenges. Vignetting correction (VC) had negligible impact since detection uses relative local background to compensate for vignetting. Geometric undistortion (GU) enabled the full pipeline to achieve near-perfect performance on larger defect sizes.

Uncertainty (Noise) Model. In practice, ground truth is unavailable—defect pixel number, location, and contrast are unknown. A high detection threshold D_{thr} (near 1) risks excessive false positives, while a low threshold may miss low-contrast defects. Our uncertainty map can be used to estimate a false positive count N_{fp} , offering a reliability metric for detection outcomes.

Since M_p already contains noise, we use M_b (computed with a larger box filter and thus assumed to be less affected by noise) for false positive estimation. We consider only pixels with $a_d = 1$ and estimate the probability that each display pixel s with $\mu(s) = M_b(s), \sigma^2(s) = \frac{\sum_c \sigma_c^2(\hat{I}_c^P(s))}{9}$ satisfies condition $M'_p(s) < M_b(s)D_{\text{thr}}$ (false positive). Note that $M'_p(s)$ is a sample drawn from normal distribution with $\mu(s)$ and $\sigma^2(s)$, and is independent of $M_p(s)$. Then,

for pixel \mathbf{s} , the z-score is given by:

$$z(\mathbf{s}) = \frac{(D_{\text{thr}} - 1) \mathbf{M}_b(\mathbf{s})}{\sigma(\mathbf{s})} \quad (24)$$

We can then compute the cumulative distribution function (CDF), representing the probability that the pixel becomes a false positive:

$$\Phi(z(\mathbf{s})) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{z(\mathbf{s})}{\sqrt{2}} \right) \right), \quad (25)$$

where erf is the Gauss error function. The final estimate of the number of false positive detections is:

$$N_{\text{fp}} = \sum_{\mathbf{s}} \Phi(z(\mathbf{s})), \quad (26)$$

Experimental results are shown in Figure 9. Without any ground truth, our method accurately predicts the number of false positives and, therefore, can be used to select the desired detection threshold.

4.4 Display color fringing

RGB subpixel layouts often introduce color fringing artifacts like red-blue edges around white text. This section evaluates whether CameraVDP can accurately predict the visibility of such artifacts.

Stimulus. To accurately reproduce color fringing, we displayed the letters “C” and “Y” on the Eizo display and captured them with the Sony $\alpha 7R$ III and FE 2.8/90 lens. The experiment included a test image (with color fringing) and a reference (without fringing). Because the pixel density of our display was too high to see color fringing, we simulated a lower resolution display (e.g., VR headset) by enlarging the captured image 4 \times in both dimensions. The reference image was generated by converting the test image to grayscale. Test and reference images are shown in Figure 7.

Experimental procedure. We investigated the effect of viewing distance on color fringing visibility using a two-alternative forced choice (2AFC) protocol. Stimuli were displayed on the Eizo monitor mounted on a motorized movable rail [Anonymous 2025]. In each trial, the display was moved to one of five viewing distances [1.00, 1.14, 1.33, 1.60, 2.00] meters, and the corresponding pixel-per-degree resolutions (with the simulated 4 \times 4 pixels) are [28.9, 32.8, 38.0, 45.4, 56.5]. Each condition was presented 20 times per observer in random order (of distances and “C”/“Y” letters). Each trial consisted of two intervals, one with color fringing (test) and one without (reference), each lasting 2 s, separated by 0.5 s of random noise. Observers were asked to select the interval containing color fringing.

Participants. We recruited 8 participants (5 male, 3 female) aged 20–35, all of whom passed a color vision and acuity screening. The experiment was approved by the departmental ethics committee.

Results and discussion. The results are shown in Figure 10, including both human data and model predictions. To account for display characteristics, the test and reference patterns were re-captured using the Sony $\alpha 7R$ III camera with the FE 1.8/35 lens before being processed by the visual difference predictor to obtain JOD values, which were then converted to detection probabilities². These constitute the model predictions, shown as purple and green lines (mean)

²Since the VDP was trained on full-field images, but color fringing stimuli are very small, JOD values were rescaled as $\text{JOD}_{\text{scale}} = \min(11.35 - 20 \times (10 - \text{JOD}), 10)$.

and shaded areas (uncertainty) in Figure 10. Human observer data are black lines in the right panel. The results demonstrate that our pipeline accurately predicts human detection probabilities of color fringing artifacts, with model uncertainty encompassing the 95% confidence interval of human results.

4.5 Display non-uniformity assessment

Section 4.3 examined high-frequency non-uniformities (defective pixels), whereas low-frequency artifacts (e.g., center-bright vignetting from backlight unevenness) are also common. This section evaluates whether CameraVDP can accurately predict their visibility as a function of viewing distance.

Stimulus. We pre-measured the Eizo display’s non-uniformity $\frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}} = 1.94\%$, which was sufficient for the experiment. The test stimulus was a 2D elliptical Gaussian with a contrast of 0.1 and a mean of 142.5 cd/m², with horizontal and vertical standard deviations set to half of the width and height, as shown in Figure 8. The reference was a uniform field matched in mean luminance. The display was driven by a 12-bit per color channel signal (10-bit native + 2 bits simulated with spatio-temporal dithering) to avoid banding.

Experimental procedure and participants. The experimental procedure and participants were nearly identical to those in Section 4.4, with two exceptions: (1) viewing distances of [0.50, 0.62, 0.80, 1.14, 2.00] meters were used; (2) to prevent detection based on luminance differences rather than non-uniformity, the reference luminance was randomly scaled by a factor in the range from 0.8 to 1.2.

Results and discussion. The results are presented in Figure 11. Overall, the findings indicate that display non-uniformities are generally perceived more strongly at greater viewing distances, as the increase in viewing distance corresponds to higher spatial frequencies, to which the human visual system is more sensitive (see Figure 8; this can also be observed by zooming the page in and out). CameraVDP accurately predicted this effect.

5 CONCLUSIONS

A consumer-grade camera can become an accurate measurement instrument when properly calibrated. Here, we calibrate the camera specifically for the task of capturing a display, so that we can take advantage of the pixel grid to refine geometric calibration and ensure accurate color measurement. Because not every display artifact is going to be visible, we integrate our camera correction pipeline with a visual difference predictor. Finally, our measurements come with an estimate of uncertainty, which is introduced by sensor noise and VDPs prediction error. Our measurement technique can be used across a range of (computational) display applications, from detecting defects to evaluating the visibility of display artifacts.

Limitations. CameraVDP has three main limitations: (1) it is applicable only to cameras that are good enough to be calibratable; (2) it requires cameras with pixel-shift capability, which does not rely on demosaicing for color reconstruction. Demosaicing may introduce inter-pixel dependencies that violate the assumption of statistical independence; (2) the increase of uncertainty due to the MTF inversion is approximated assuming white noise, which may not hold for scenes with large variation of intensity.

REFERENCES

- Cecilia Aguerrebere, Julie Delon, Yann Gousseau, and Pablo Musé. 2014. Best algorithms for HDR image generation. A study of performance bounds. *SIAM Journal on Imaging Sciences* 7, 1 (2014), 1–34.
- A. Anonymous. 2025. Withheld for anonymity reasons. 1, 1 (2025), 1–10.
- Philipp Arras, Philipp Frank, Reimar Leike, Rüdiger Westermann, and Torsten A Enßlin. 2019. Unified radio interferometric calibration and imaging with joint uncertainty quantification. *Astronomy & Astrophysics* 627 (2019), A134.
- Naoki Asada, Akira Amano, and Masashi Baba. 1996. Photometric calibration of zoom lens systems. In *Proceedings of 13th International Conference on Pattern Recognition*, Vol. 1. IEEE, 186–190.
- Maliha Ashraf, Rafal K Mantiuk, Alexandre Chapiro, and Sophie Wuerger. 2024. castleCSF—A contrast sensitivity function of color, area, spatiotemporal frequency, luminance and eccentricity. *Journal of vision* 24, 4 (2024), 5–5.
- Johnathan M Bardsley. 2012. MCMC-based image reconstruction with uncertainty quantification. *SIAM Journal on Scientific Computing* 34, 3 (2012), A1316–A1332.
- Peter GJ Barten. 2003. Formula for the contrast sensitivity of the human eye. In *Image Quality and System Performance*, Vol. 5294. SPIE, 231–238.
- MJ Bayarri, JO Berger, and F Liu. 2009. Modularization in Bayesian analysis, with emphasis on analysis of computer models. (2009).
- David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112, 518 (2017), 859–877.
- Avery E Broderick, Dominic W Pesce, Paul Tiede, Hung-Yi Pu, and Roman Gold. 2020. Hybrid very long baseline interferometry imaging and modeling with THEMIS. *The Astrophysical Journal* 898, 1 (2020), 9.
- Peter D Burns, Kenichiro Masaoka, Kenneth Parulski, and Dietmar Wueller. 2022. Updated camera spatial frequency response for ISO 12233. *Electronic Imaging* 34 (2022), 1–6.
- Yancheng Cai, Ali Bozorgian, Maliha Ashraf, Robert Wanat, and K Rafal Mantiuk. 2024. elatCSF: A Temporal Contrast Sensitivity Function for Flicker Detection and Modeling Variable Refresh Rate Flicker. In *SIGGRAPH Asia 2024 Conference Papers*. 1–11.
- John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.
- Scott J Daly. 1992. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*, Vol. 1666. SPIE, 2–15.
- Paul E Debevec and Jitendra Malik. 2023. Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 643–652.
- Armen Der Kiureghian and Ove Ditlevsen. 2009. Aleatory or epistemic? Does it matter? *Structural safety* 31, 2 (2009), 105–112.
- Canberk Ekmekci and Muidat Cetin. 2021. What does your computational imaging algorithm not know?: A Plug-and-Play model quantifying model uncertainty. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4018–4027.
- Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. 2015. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing* 24, 5 (2015), 1460–1470.
- Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.
- Annika Hagemann, Moritz Knorr, Holger Janssen, and Christoph Stiller. 2022. Inferring bias and uncertainty in camera calibration. *International Journal of Computer Vision* (2022), 1–16.
- Param Hanji, Fangcheng Zhong, and Rafal K Mantiuk. 2020. Noise-aware merging of high dynamic range image stacks without camera calibration. In *European Conference on Computer Vision*. Springer, 376–391.
- Sing Bing Kang and Richard Weiss. 2000. Can we calibrate a camera using an image of a flat, textureless Lambertian surface?. In *Computer Vision—ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part II* 6. Springer, 640–653.
- Marc C Kennedy and Anthony O’Hagan. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 3 (2001), 425–464.
- Sarah Kerr. 2024. Recommendations for the detection and analysis of the ISO 12233: 2023 e-SFR slanted star. *Electronic Imaging* 36 (2024), 1–6.
- Dirk P Kroese, Thomas Taimre, and Zdravko I Botev. 2013. *Handbook of monte carlo methods*. John Wiley & Sons.
- Christian Loebich, Dietmar Wueller, Bruno Kligen, and Anke Jaeger. 2007. Digital camera resolution measurements using sinusoidal Siemens stars. In *Digital Photography III*, Vol. 6502. SPIE, 214–224.
- Rafal K Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–19.
- Rafal K Mantiuk, Dounia Hammou, and Param Hanji. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625* (2023).
- Rafal K Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVDP: A visual difference predictor for image, video and display distortions. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–20.
- Juvi Mustonen, Jyrki Rovamo, and Risto Näsänen. 1993. The effects of grating area and spatial frequency on contrast sensitivity as a function of light level. *Vision research* 33, 15 (1993), 2065–2072.
- Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. 2020. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory* 1, 1 (2020), 39–56.
- Sascha Ranftl, Wolfgang von der Linden, and MaxEnt 2021 Scientific Committee. 2021. Bayesian surrogate analysis and uncertainty propagation. In *Physical Sciences Forum*, Vol. 3. MDPI, 6.
- Audrey Repetti, Marcelo Pereyra, and Yves Wiaux. 2019. Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM Journal on Imaging Sciences* 12, 1 (2019), 87–118.
- Danilo Rezende and Shakir Mohamed. 2015. Variational inference with normalizing flows. In *International conference on machine learning*. PMLR, 1530–1538.
- Jyrki Rovamo, Olavi Luntinen, and Risto Näsänen. 1993. Modelling the dependence of contrast sensitivity on grating area and spatial frequency. *Vision research* 33, 18 (1993), 2773–2788.
- Otto H Schade Sr. 1956. Optical and photoelectric analog of the eye. *Journal of the optical Society of America* 46, 9 (1956), 721–739.
- He Sun and Katherine L Bouman. 2021. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 2628–2637.
- Zhongwei Tang, Rafael Grompone Von Gioi, Pascal Monasse, and Jean-Michel Morel. 2017. A precision analysis of camera distortion models. *IEEE Transactions on Image Processing* 26, 6 (2017), 2694–2704.
- Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-François Lalonde, and Felix Heide. 2021. Differentiable Compound Optics and Processing Pipeline Optimization for End-to-end Camera Design. *ACM Transactions on Graphics* 40, 2 (April 2021), 1–19. <https://doi.org/10.1145/3446791>
- Andrew B Watson. 1993. DCTune: A technique for visual optimization of DCT quantization matrices for individual images. In *Sid International Symposium Digest of Technical Papers*, Vol. 24. Society for information display, 946–946.
- Andrew B Watson et al. 1986. Temporal sensitivity. *Handbook of perception and human performance* 1, 6 (1986), 1–43.
- Juyang Weng, Paul Cohen, Marc Herniou, et al. 1992. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence* 14, 10 (1992), 965–980.
- Yujia Xue, Shiyi Cheng, Yunzhe Li, and Lei Tian. 2019. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica* 6, 5 (2019), 618–629.
- Yuanjie Zheng, Stephen Lin, Chandra Kambhampettu, Jingyi Yu, and Sing Bing Kang. 2008. Single-image vignetting correction. *IEEE transactions on pattern analysis and machine intelligence* 31, 12 (2008), 2243–2256.

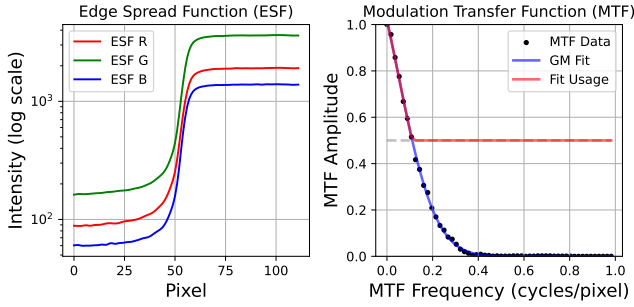


Fig. 4. Edge Spread Function (ESF) and Modulation Transfer Function (MTF) results. To prevent excessive amplification of noise, the MTF is constrained to a minimum value of 0.5 (indicated by the red line). Images were captured of the diagonal black-white edge at the center of a Siemens star rotated by 45° .

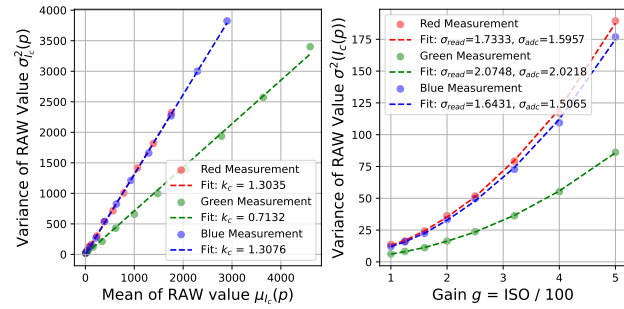


Fig. 5. Noise model parameter fitting (using Sony FE 90/2.8 lens). Left: linear fitting in bright scenes; right: quadratic fitting in dark scenes. Note that green channels are averaged across two measurements in 4-shot pixel-shift capture, so their variance is halved compared to red and blue.

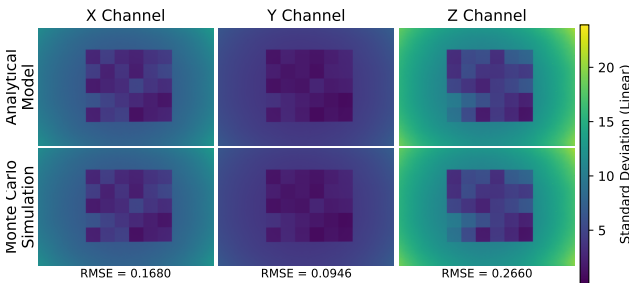


Fig. 6. Monte Carlo simulations validate the analytical model of uncertainty. The top row shows results from the theoretical derivation, while the bottom row shows results of 1000 Monte Carlo samples.

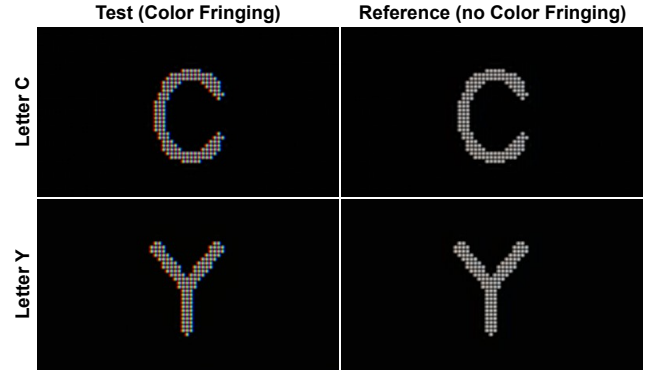


Fig. 7. Test and reference images for color fringing psychophysical experiments. The images have been center-cropped to $10\times$ magnification to facilitate comparison; the actual stimuli used in the experiments were significantly smaller. Note that due to variations in display primaries, the test (left) and reference (right) images may appear chromatically different on some displays, although they are perceptually matched on the Eizo display.

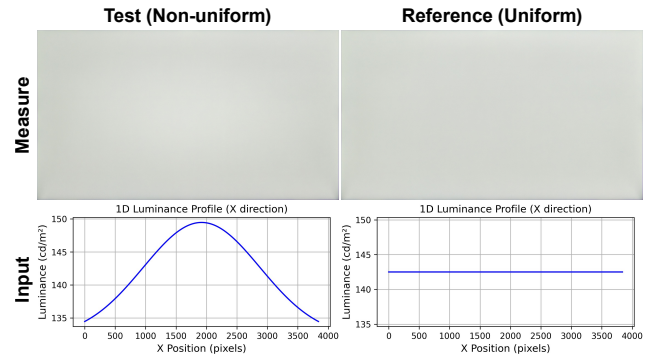


Fig. 8. Test and reference images for display uniformity assessment psychophysical experiments. The first row shows the display output measured by Sony FE 1.8/35 (converted to sRGB), and the second row shows the 1D luminance profile of the input signal along the X-axis. Reducing the page scale will make the central bright spot in the test stimulus more prominent.

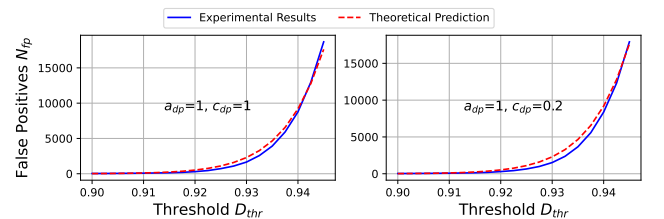


Fig. 9. Estimating false positive number via uncertainty map for trustworthy detection without any prior knowledge of defective pixel conditions.; blue: empirical, red dashed: theoretical from uncertainty; left/right: different contrast corruptions (contrast does not affect false positive number). Only $D_{thr} \in [0.900, 0.945]$ is shown—smaller values yield near-zero false positives, while larger values exceed the computational limits of NMS.

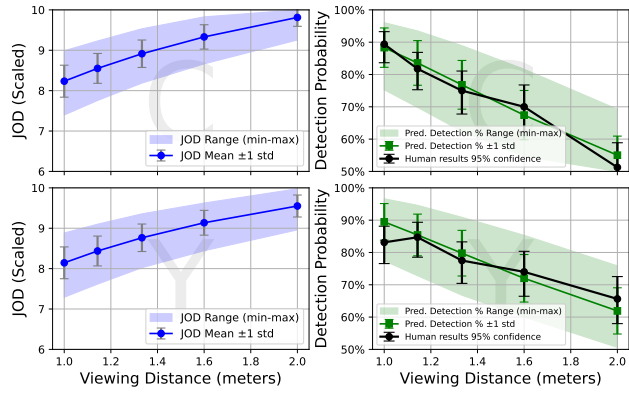


Fig. 10. Color fringing experiment results. Top: letter C; bottom: letter Y. Left: the scaled JOD values predicted by the visual difference predictor, along with standard deviation (error bars) and extrema (shader region) as measures of uncertainty. Right: the human experimental results (mean and 95% confidence intervals) in black, and the predicted detection probability (mean, standard deviation, and extrema converted from JOD). The predictions closely match human performance.

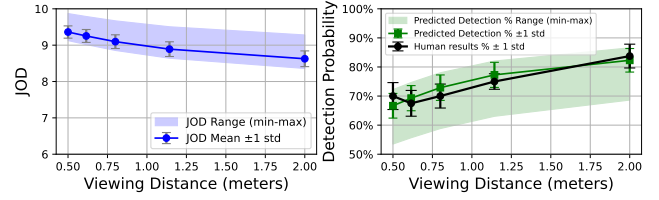


Fig. 11. Display non-uniformity assessment experiment results. The notation is the same as in Figure 10.

CameraVDP: Perceptual Display Assessment with Uncertainty Estimation via Camera and Visual Difference Prediction - Supplementary Materials

ANONYMOUS AUTHOR(S)

SUBMISSION ID: 1077

This supplementary material includes: (1) a table of all variables and their definitions, (2) a complete derivation of uncertainty propagation in MTF inversion, and (3) the computation of mean and uncertainty in color correction for multi-primary displays (e.g., WRGB OLEDs).

ACM Reference Format:

Anonymous Author(s). 2025. CameraVDP: Perceptual Display Assessment with Uncertainty Estimation via Camera and Visual Difference Prediction - Supplementary Materials. In . ACM, New York, NY, USA, 3 pages. <https://doi.org/XXXXXXX.XXXXXXX>

CONTENT

- Section S1 — Uncertainty propagation in MTF inversion
- Section S2 — Color Correction for 4+ primaries
- Table S1 — List of symbols
- Table S2 — Fit parameters

S1 UNCERTAINTY PROPAGATION IN MTF INVERSION

In the main text, we provide only a brief derivation and result of the MTF inversion. Here, we present the complete derivation and proof.

The scene radiance $\Psi_c(\mathbf{p})$ affected by blur and glare can be modeled as the convolution of the clean scene radiance $\mathbf{L}_c^M(\mathbf{p})$ and the camera point spread function (PSF) $P(\mathbf{p})$, plus noise $\eta(\mathbf{p})$:

$$\Psi_c(\mathbf{p}) = (\mathbf{L}_c^M * P)(\mathbf{p}) + \eta(\mathbf{p}), \quad (\text{S1})$$

where $*$ denotes convolution. The MTF $M(\omega)$ is the modulus of the Fourier transform of the PSF $P(\mathbf{p})$ [Burns et al. 2022], ω is spatial frequency in cycles-per-pixel. Taking the Fourier transform \mathcal{F} on both sides of Eq. (S1) yields:

$$\mathcal{F}(\Psi_c) = \mathcal{F}(\mathbf{L}_c^M)M(\omega) + \eta'(\omega), \quad (\text{S2})$$

Considering the noise, we use Wiener deconvolution to obtain the degraded and deblurred estimate of \mathbf{L}_c^M :

$$\hat{\mathbf{L}}_c^M(\mathbf{p}) = \mathcal{F}^{-1} \left(\mathcal{F}(\hat{\Psi}_c) G_c(\omega) \right) (\mathbf{p}), \quad (\text{S3})$$

where the Wiener filter is ($*$ is the conjugate operator):

$$G_c(\omega) = \frac{M^*(\omega) S_{\hat{\Psi}_c}(\omega)}{|M(\omega)|^2 S_{\hat{\Psi}_c}(\omega) + N_{\hat{\Psi}_c}(\omega)}. \quad (\text{S4})$$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGGRAPH 2025, Conference Location,

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/XXXXXXX.XXXXXXX>

$S_{\hat{\Psi}_c}(\omega)$ and $N_{\hat{\Psi}_c}(\omega)$ represent the power spectral densities (PSD) of the signal $\mathbb{S}_{\hat{\Psi}_c}(\mathbf{p})$ and noise $\mathbb{N}_{\hat{\Psi}_c}(\mathbf{p})$:

$$\hat{\Psi}_c(\mathbf{p}) = \mathbb{S}_{\hat{\Psi}_c}(\mathbf{p}) + \mathbb{N}_{\hat{\Psi}_c}(\mathbf{p}), \quad (\text{S5})$$

$$\mathbb{S}_{\hat{\Psi}_c}(\mathbf{p}) \sim N \left(\mu_{\hat{\Psi}_c}(\mathbf{p}), 0 \right), \quad (\text{S6})$$

$$\mathbb{N}_{\hat{\Psi}_c}(\mathbf{p}) \sim N \left(0, \sigma_{\hat{\Psi}_c}^2(\mathbf{p}) \right). \quad (\text{S7})$$

Thus $S_{\hat{\Psi}_c}(\omega)$ can be directly computed:

$$S_{\hat{\Psi}_c}(\omega) = \left| \mathcal{F} \left(\mathbb{S}_{\hat{\Psi}_c} \right) \right|^2 = \left| \mathcal{F}(\mu_{\Psi_c}) \right|^2. \quad (\text{S8})$$

However, since $\mathbb{N}_{\hat{\Psi}_c}(\mathbf{p})$ is a random variable and not directly accessible, the computation of $N_{\hat{\Psi}_c}(\omega)$ differs accordingly. The noise $\mathbb{N}_{\hat{\Psi}_c}(\mathbf{p})$ is assumed to be white in the spatial frequency domain (i.e., equal magnitude in all frequencies). According to Parseval's theorem, the energy of the noise signal in the spatial domain equals that in the frequency domain, so:

$$N_{\hat{\Psi}_c}(\omega) \approx \overline{\sigma_{\hat{\Psi}_c}^2(\mathbf{p})}. \quad (\text{S9})$$

For simplicity, the MTF was assumed to be isotropic. We measured $M(\omega)$ using the slanted-edge method, as shown in the second column of Figure M2 (main text). Specifically, images were captured of the diagonal black-white edge at the center of a Siemens star rotated by 45°. The MTF was fitted using:

$$M'(\omega) = a_1 \exp \left(- \left(\frac{\omega - b_1}{c_1} \right)^2 \right) + a_2 \exp \left(- \left(\frac{\omega - b_2}{c_2} \right)^2 \right), \quad (\text{S10})$$

where $a_1, a_2, b_1, b_2, c_1, c_2$ are fitting parameters. To suppress noise amplification, $M(\omega) = \max(M'(\omega), 0.5)$, as shown in Figure M4.

Since the Fourier transform is a linear operation, $\hat{L}_c(\mathbf{p})$ follows a normal distribution. The mean of $\hat{\mathbf{L}}_c^M(\mathbf{p})$ is given by Eq. (S3):

$$\mu_{\hat{\mathbf{L}}_c^M}(\mathbf{p}) = \mathcal{F}^{-1} \left(\mathcal{F}(\mu_{\Psi_c}) G(\omega) \right) (\mathbf{p}), \quad (\text{S11})$$

Now we aim to compute the variance $\sigma^2(\hat{\mathbf{L}}_c^M(\mathbf{p}))$. Our analysis relies on the following assumptions: (1) the camera PSF P decays rapidly in spatial domain; (2) the noise follows a white noise distribution; and (3) local variance is approximately uniform within small pixel neighborhoods. These assumptions hold for most well-focused cameras and typical scenes.

Revisiting Eq. (S1), since P decays rapidly in spatial domain, the following expression still holds within the neighborhood of pixel \mathbf{p} :

$$\Psi'_c(\mathbf{p}) = (\mathbf{L}_c'^M * P)(\mathbf{p}) + \eta(\mathbf{p}), \quad (\text{S12})$$

where $'$ indicates values within the neighborhood of pixel \mathbf{p} .

From Eq. (S3), we have:

$$\mathcal{F}(\hat{\mathbf{L}}_c'^M)(\omega) = \left(\mathcal{F}(\hat{\Psi}_c') G \right) (\omega). \quad (\text{S13})$$

Table S1. Table of key symbol definitions used in the main text.

Variable Symbols	Definition
a_d	Side length of the simulated defective pixel in the defective pixel detection experiment, expressed in units of display pixel size.
c	Color channels in the camera RAW image, $c \in \{r, g, b\}$.
c_d	Weber contrast of the simulated defective pixel.
C_c	The spectral sensitivity of camera channel $c \in \{r, g, b\}$.
d	Camera dark noise.
D_{thr}	Defective pixel detection threshold, $D_{thr} \in [0, 1]$.
E_k	Spectral power distribution (SPD) of display subpixels, $k \in \{R, G, B\}$.
\mathcal{F}	Fourier transform.
g	Camera gain.
\mathbf{g}	Camera pixel coordinates in undistorted images, with the image center as the origin; x-axis pointing right, y-axis pointing upward.
G	Wiener filter.
h	Coordinate mapping of the homography transformation, $h : \mathbf{s} \rightarrow \mathbf{g}$.
\mathbf{H}	Homography matrix.
i	Each RAW exposure in HDR image stack.
\mathbf{I}	RAW (digital) sensor values captured by the camera, represented as a matrix of linear values.
J	Jacobian matrix used in the uncertainty propagation for display with more than three subpixels.
k	Display subpixels, $k \in \{R, G, B\}$.
k_c	Camera quantum efficiency of camera color channels.
\mathbf{L}^M	Clean scene radiance (result of MTF inversion).
\mathbf{L}^V	Result of vignetting correction.
\mathbf{L}^D	Result of geometric transformation.
ℓ	Log-likelihood function.
m	Global coordinate mapping of the geometric transformation, $m : \mathbf{s} \rightarrow \mathbf{p}$.
$M(\omega)$	Modulation transfer function (MTF).
\mathbf{M}_b	Background mean map in the defective pixel detection experiment.
\mathbf{M}_p	Mean (pixel surrounding) map in the defective pixel detection experiment.
\mathbf{M}	Color correction matrix for display with three subpixels.
\mathcal{M}	Color correction matrix for display with more than three subpixels.
N	Total number of RAW images in the HDR image stack.
$N_{(\cdot)}(\omega)$	The power spectral densities (PSD) of the noise.
$\mathbb{N}_{(\cdot)}(\mathbf{p})$	Noise component of the variable (zero mean).
o	Display pixel oversampling factor.
\mathbf{p}	Camera pixel coordinates in RAW images, with the image center as the origin; x-axis pointing right, y-axis pointing upward.
P_k	Linearized display subpixel value, $k \in \{R, G, B\}$.
P	Camera point spread function (PSF).
\mathbf{s}	Display pixel coordinates in Display panel, with the display center as the origin; x-axis pointing right, y-axis pointing upward.
$S_{(\cdot)}(\omega)$	The power spectral densities (PSD) of the signal.
$\mathbb{S}_{(\cdot)}(\mathbf{p})$	Signal component of the variable (zero variance).
\bar{S}	The CIE 1931 standard observer color matching functions (CMFs).
t	Camera exposure time, in seconds (s).
u	Coordinate mapping of the geometric undistortion, $u : \mathbf{g} \rightarrow \mathbf{p}$.
V	Vignetting function.
x	A single measurement of \mathbf{X} .
\mathbf{X}	Relative radiance, an intermediate variable for computing Ψ .
Y	Result of color correction, in XYZ color space.
ϵ	Noise in Vignetting correction.
η	Noise in MTF inversion.
λ	The wavelength.
$\mu_{(\cdot)}$	Mean of the normal distribution.
$\sigma_{(\cdot)}^2$	Variance of the normal distribution.
$\Sigma_{(\cdot)}$	Covariance matrix of the multivariate normal distribution.
σ_{adc}^2	Camera analog-to-digital conversion noise (ADC noise).
σ_{read}^2	Camera read noise.
Ψ	Scene radiance, the reconstruction target in the HDR image stack merging section.
ω	Spatial frequency in cycles-per-pixel.
Pois	Poisson distribution.
\mathcal{N}	Normal distribution.

Table S2. The fitted parameters.

Part	Parameters
Noise Model	$k_c = [1.303514, 0.713188, 1.307612],$ $\sigma_{\text{read}}^2 = [1.733335, 2.074783, 1.643126],$ $\sigma_{\text{adc}}^2 = [1.595734, 2.021769, 1.506513]$
MTF	$a_1 = 0.00174, b_1 = 0.67193, c_1 = 0.12362,$ $a_2 = 1.30353, b_2 = -0.11405, c_2 = 0.22962$

Decompose each component into the sum of signal and noise:

$$\mathcal{F} \left(\hat{\mathbf{S}}_{\hat{\mathbf{L}}_c^M} + \mathbf{N}_{\hat{\mathbf{L}}_c^M} \right) (\omega) = \left(\mathcal{F} \left(\hat{\mathbf{S}}_{\hat{\Psi}_c'} + \mathbf{N}_{\hat{\Psi}_c'} \right) G \right) (\omega). \quad (\text{S14})$$

Due to the linearity of the Fourier transform:

$$\mathcal{F} \left(\mathbf{N}_{\hat{\mathbf{L}}_c^M} \right) (\omega) = \left(\mathcal{F} \left(\mathbf{N}_{\hat{\Psi}_c'} \right) G \right) (\omega). \quad (\text{S15})$$

Therefore, the PSD of the noise components of $\hat{\mathbf{L}}_c^M$ are:

$$\begin{aligned} N_{\hat{\mathbf{L}}_c^M}(\omega) &= \left| \mathcal{F} \left(\mathbf{N}_{\hat{\mathbf{L}}_c^M} \right) (\omega) \right|^2 = \left| \mathcal{F} \left(\mathbf{N}_{\hat{\Psi}_c'} \right) (\omega) G(\omega) \right|^2 \\ &= \left| \mathcal{F} \left(\mathbf{N}_{\hat{\Psi}_c'} \right) (\omega) \right|^2 |G(\omega)|^2 = N_{\hat{\Psi}_c'}(\omega) |G(\omega)|^2. \end{aligned} \quad (\text{S16})$$

By the Wiener-Khinchin theorem, the autocorrelation function $R(\tau_x, \tau_y)$ (τ_x, τ_y are spatial shifts) of a wide-sense stationary random process (assumption (3)) and its PSD $N(\omega)$ form a Fourier transform pair:

$$R(\tau_x, \tau_y) = \mathcal{F}^{-1}(N(\omega)). \quad (\text{S17})$$

Since the noise \mathbf{N} is a zero-mean stochastic process, we have:

$$\sigma_{\hat{\mathbf{L}}_c^M}^2(\mathbf{p}) \approx R_{\hat{\mathbf{L}}_c^M}(0, 0) = \iint N_{\hat{\mathbf{L}}_c^M}(\omega) d\omega, \quad (\text{S18})$$

where second equals sign follows the definition of the inverse Fourier transform. Under the $\hat{\Psi}_c$ white noise assumption:

$$N_{\hat{\Psi}_c'}(\omega) \approx \sigma_{\hat{\Psi}_c'}^2(\mathbf{p}), \quad (\text{S19})$$

considering Eq. (S16) and Eq. (S18), we have:

$$\sigma_{\hat{\mathbf{L}}_c^M}^2(\mathbf{p}) \approx \iint N_{\hat{\mathbf{L}}_c^M}(\omega) d\omega \approx \sigma_{\hat{\Psi}_c'}^2(\mathbf{p}) \iint |G(\omega)|^2 d\omega. \quad (\text{S20})$$

The \approx is used because the derivation relies on the three aforementioned assumptions. The accuracy is validated via Monte Carlo simulations in Section M4.2 in main text.

S2 COLOR CORRECTION FOR 4+ PRIMARIES

We derived the color correction method for three-primary displays (RGB subpixels) in the main text. However, some advanced displays aiming for HDR and wide color gamut utilize more than three primaries—for example, OLED displays include an additional white subpixel. This section presents the derivation of the mean and variance for color correction in displays with four or more subpixels.

The radiance $\hat{\mathbf{L}}_c^D$ recorded by the camera for a single display pixel is:

$$\hat{\mathbf{L}}_c^D = \sum_k P_k \int_{\lambda} E_k(\lambda) C_c(\lambda) d\lambda, \quad (\text{S21})$$

where λ is the wavelength, $E_k(\lambda)$ is the spectral power distribution (SPD) of subpixel $k \in \{R, G, B\}$ at unit intensity, P_k is the linearized subpixel value, and $C_c(\lambda)$ denotes the spectral sensitivity of camera channel $c \in \{r, g, b\}$.

Similarly, the trichromatic CIE XYZ value (related to cone responses) can be expressed as:

$$\mathbf{Y} = \sum_k P_k \int_{\lambda} E_k(\lambda) \tilde{\mathbf{S}}(\lambda) d\lambda, \quad (\text{S22})$$

where $\mathbf{Y} = [X, Y, Z]'$ and $\tilde{\mathbf{S}}(\lambda) = [\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)]^T$ are the CIE 1931 standard observer color matching functions (CMFs).

For a display with three color subpixels, there exists a unique transformation matrix $\mathbf{M} \in \mathbb{R}^{3 \times 3}$ such that:

$$\hat{\mathbf{Y}}(\mathbf{s}) = \mathbf{M} \hat{\mathbf{L}}^D(\mathbf{s}). \quad (\text{S23})$$

However, advanced OLED displays employ four subpixels (R, G, B, W) with complex driving mechanisms [Ashraf et al. 2024], allowing only an approximate transformation matrix to be derived. To improve color correction accuracy, we employed the root polynomial regression method [Finlayson et al. 2015], which extends \mathbf{L}^D to

$$\mathcal{L}^D(\mathbf{s}) = \left[\mathbf{L}_r^D, \mathbf{L}_g^D, \mathbf{L}_b^D, \sqrt{\mathbf{L}_r^D \mathbf{L}_g^D}, \sqrt{\mathbf{L}_r^D \mathbf{L}_b^D}, \sqrt{\mathbf{L}_g^D \mathbf{L}_b^D} \right]^T(\mathbf{s}), \quad (\text{S24})$$

and fitted a transformation matrix $\mathcal{M} \in \mathbb{R}^{3 \times 6}$ such that:

$$\hat{\mathbf{Y}}(\mathbf{s}) = \mathcal{M} \mathcal{L}^D(\mathbf{s}), \quad \mu_{\hat{\mathbf{Y}}}(\mathbf{s}) = \mathcal{M} \mu_{\mathcal{L}^D}(\mathbf{s}). \quad (\text{S25})$$

To estimate uncertainty, we need to model covariance as the color channels are no longer independent. For the display with more than three primaries, we have used root polynomial regression, and thus apply a first-order Taylor approximation (i.e., Jacobian matrix method) to propagate uncertainty. The Jacobian matrix $J \in \mathbb{R}^{6 \times 3}$ obtained from the root polynomial regression method is:

$$J = \frac{\partial \mathcal{L}^D(\mathbf{s})}{\partial (\mathbf{L}_r^D(\mathbf{s}), \mathbf{L}_g^D(\mathbf{s}), \mathbf{L}_b^D(\mathbf{s}))}, \quad J_{ij} = \frac{\partial (\mathcal{L}^D)_i(\mathbf{s})}{\partial (\mathbf{L}^D)_j(\mathbf{s})}. \quad (\text{S26})$$

The covariance matrix of \mathcal{L}^D and color-corrected result \mathbf{Y} are:

$$\Sigma_{\hat{\mathbf{L}}^D}(\mathbf{s}) = J \Sigma_{\mathcal{L}^D}(\mathbf{s}) J^T \quad (\text{S27})$$

$$\Sigma_{\hat{\mathbf{Y}}}(\mathbf{s}) = \mathcal{M} \Sigma_{\hat{\mathbf{L}}^D}(\mathbf{s}) \mathcal{M}^T = \mathcal{M} J \Sigma_{\mathcal{L}^D}(\mathbf{s}) J^T \mathcal{M}^T, \quad (\text{S28})$$

where $\Sigma_{(\cdot)}$ is the covariance matrix.

REFERENCES

- Maliha Ashraf, Alejandro Sztrajman, Dounia Hammou, and Rafal K Mantiuk. 2024. Color calibration methods for OLED displays. *Electronic Imaging* 36 (2024), 1–7.
- Peter D Burns, Kenichiro Masaoka, Kenneth Parulski, and Dietmar Wueller. 2022. Updated camera spatial frequency response for ISO 12233. *Electronic Imaging* 34 (2022), 1–6.
- Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. 2015. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing* 24, 5 (2015), 1460–1470.

¹Note that color correction is performed on a per-captured-image-pixel basis.