

Towards Imperceptible Watermarking Via Environment Illumination for Consumer Cameras

Hodaka Kawachi[†] Tomoya Nakamura[†] Hiroaki Santo[†] SaiKiran Kumar Tedla[‡]
Trevor D. Canham[‡] Yasushi Yagi[†] Michael S. Brown[‡]

[†]The University of Osaka [‡]York University

Abstract

This paper introduces a method for using LED-based environmental lighting to produce visually imperceptible watermarks for consumer cameras. Our approach optimizes an LED light source’s spectral profile to be minimally visible to the human eye while remaining highly detectable by typical consumer cameras. The method jointly considers the human visual system’s sensitivity to visible spectra, modern consumer camera sensors’ spectral sensitivity, and narrowband LEDs’ ability to generate broadband spectra perceived as “white light” (specifically, D65 illumination). To ensure imperceptibility, we employ spectral modulation rather than intensity modulation. Unlike conventional visible light communication, our approach enables watermark extraction at standard low frame rates (30–60 fps). While the information transfer rate is modest—embedding 128 bits within a 10-second video clip—this capacity is sufficient for essential metadata supporting privacy protection and content verification.

1. Introduction

This paper explores a method for using LED-based environmental lighting to physically “watermark” a scene in a way that is invisible to viewers and unnoticeable in recorded video. This watermark can help prevent unauthorized recording when integrated with camera manufacturers, as cameras can be designed to detect the watermark and disable recording in restricted areas. This approach also provides a means to verify video authenticity, revealing when a scene was not captured at the intended location. This technique holds promise as a future privacy measure and a safeguard against deepfake content.

To achieve this, we propose an active illumination approach using narrowband LEDs capable of high-speed modulation. To reliably embed and extract information with consumer cameras operating at 30–60 fps, illumination must be modulated at or below half the camera’s frame rate

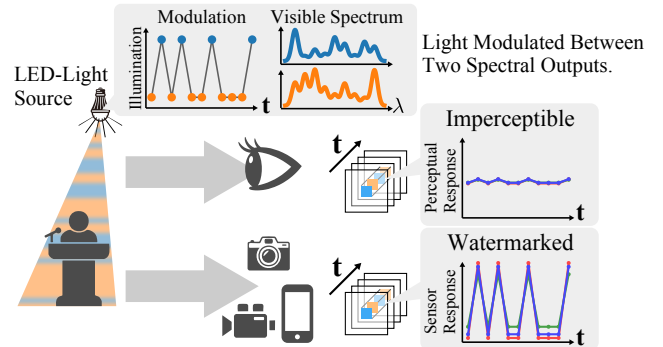


Figure 1. Overview of the proposed method. Using two optimized spectra as lighting and switching them at 15 fps, the flickering is completely imperceptible to humans but detectable by the camera, acting as a watermark.

(i.e., 15 fps), following Shannon’s sampling theorem. However, modulation within this range is highly perceptible to humans and can cause discomfort, including risks such as photosensitive epilepsy. This constraint makes direct temporal modulation impractical for non-intrusive applications.

To overcome this limitation, we employ temporal spectral modulation, leveraging differences in spectral sensitivity between CMOS sensors and the human visual system [13]. Instead of modulating light intensity, we adjust the LED spectra to meet three key criteria: (1) minimal detectability by human observers, (2) maximum detectability by CMOS sensors, and (3) a white appearance equivalent to standard D65 illumination. By exploiting this spectral sensitivity disparity, we generate optical changes imperceptible to humans yet detectable by cameras, enabling low-frequency spectral modulation without visual discomfort while preserving information embedding.

Although the system transmits data at a relatively low rate of 15 bits per second, it can embed up to 128 bits within a 10-second video clip. This capacity is sufficient to encode essential metadata, such as location information, recording permissions, and timestamps—key details that support privacy protection and content verification.

Contributions To the best of our knowledge, we present the first method for optically embedding information into LED illumination sources using spectral modulation, enabling extraction from video recordings of the scene. By leveraging an optimization approach that accounts for differences in spectral sensitivity between humans and cameras, we achieve spectral variations detectable only by cameras. We validate our method’s feasibility through a user study with a prototype lighting system and various consumer cameras, demonstrating detection from videos captured across different devices and through unknown in-camera processing, while confirming that the signal remains below the human visibility threshold. Beyond watermarking, our approach may serve as a human-invisible, camera-centric active cue for broader vision tasks—for example, near-infrared “dark-flash” illumination that aids geometry/appearance estimation [29]—with related explorations indicating similar potential for depth sensing and for hyperspectral (and HS+3D) reconstruction using commodity RGB cameras [20, 22, 23, 31].

2. Related Work

This section reviews related work in visible light communication (VLC), spatial watermarking, hyperspectral lighting, and the characteristics of human and camera color perception.

Visible Light Communication (VLC). VLC is a wireless communication technology that transmits data using visible light, typically emitted by LEDs [11, 15, 16, 21, 26, 27]. Unlike radio-frequency (RF) communication, VLC exploits the rapid on-off switching capabilities of LEDs, allowing them to function as both illumination and data-transmitting sources. This high-frequency flickering, imperceptible to the human eye, makes VLC promising for applications such as indoor positioning and secure data transfer in RF-sensitive environments (e.g., hospitals and airplanes).

To achieve VLC with image sensors, the low frame rate of typical cameras necessitates data transmission at relatively low frequencies near the human perceptual threshold (30–60 Hz). Several studies have explored rolling shutter image sensors, which can capture high-frequency flickering signals due to the sequential exposure of pixels across the sensor [5, 12, 19]. While this approach enables high-frequency VLC with cameras, it is impractical for one-to-many communication because variations in shutter speed and exposure time require calibration or prior knowledge of each camera’s settings. Thus, to ensure universal compatibility, it is crucial to use low-frequency, imperceptible lighting that remains undetectable to the human eye yet reliably detectable by camera sensors.

Spatial Embedding. Recent CVPR/ICCV work has rapidly advanced invisible digital watermarking for validating gen-

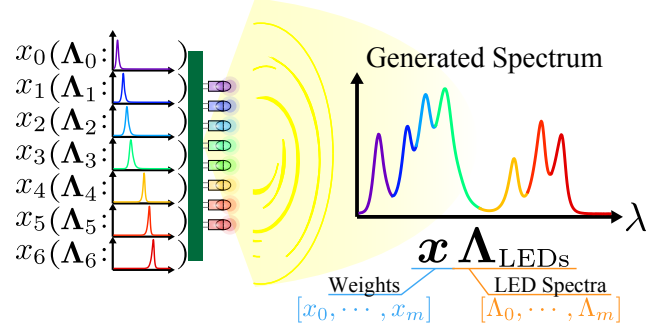


Figure 2. Configuration of a hyperspectral lighting system using narrowband LEDs. By controlling the emission intensity of each LED, a custom spectral output is generated through the weighted sum of individual LED spectra.

erative content [4, 30]. Similarly, in the spatial domain, information can be embedded by projecting patterns using structured light or projection mapping [2, 8]. Although these spatial methods can embed more information through spatial addressing and multiplexing, they usually require pre-mapping the scene and precise alignment between the projector and surfaces, which makes them difficult to deploy and maintain in large or changing environments. In addition, making projected cues truly imperceptible in everyday settings remains challenging.

Hyperspectral Illumination. Recent advances in narrowband LED technology have enabled the construction of hyperspectral lighting systems by combining multiple LEDs [32]. As shown in Fig. 2, the spectrum of a narrowband LED lighting system can be customized by adjusting the emission intensities of individual LEDs, which combine as a weighted sum of each LED’s spectral profile. Traditionally, controllable hyperspectral lighting required specialized light sources, typically involving diffraction gratings or spatial light modulators [33]. In contrast, the multi-LED approach offers a simpler and more accessible way to design lighting with tailored spectral characteristics, enabling customized spectral profiles.

Human Color Perception and Metamerism. Human color perception is governed by the trichromatic theory, which states that color vision relies on three types of photoreceptor cells in the retina, each sensitive to different regions of the visible spectrum. These sensitivities, often represented by the CIE XYZ color matching functions [6, 28], serve as the foundation for standard color spaces that model human vision [1, 10]. However, metamerism allows different spectral power distributions (SPDs) to produce identical color sensations for human observers [7, 25]. This occurs because the brain interprets color based on the combined responses of the three photoreceptor types, meaning distinct light spectra can elicit the same perceptual response.

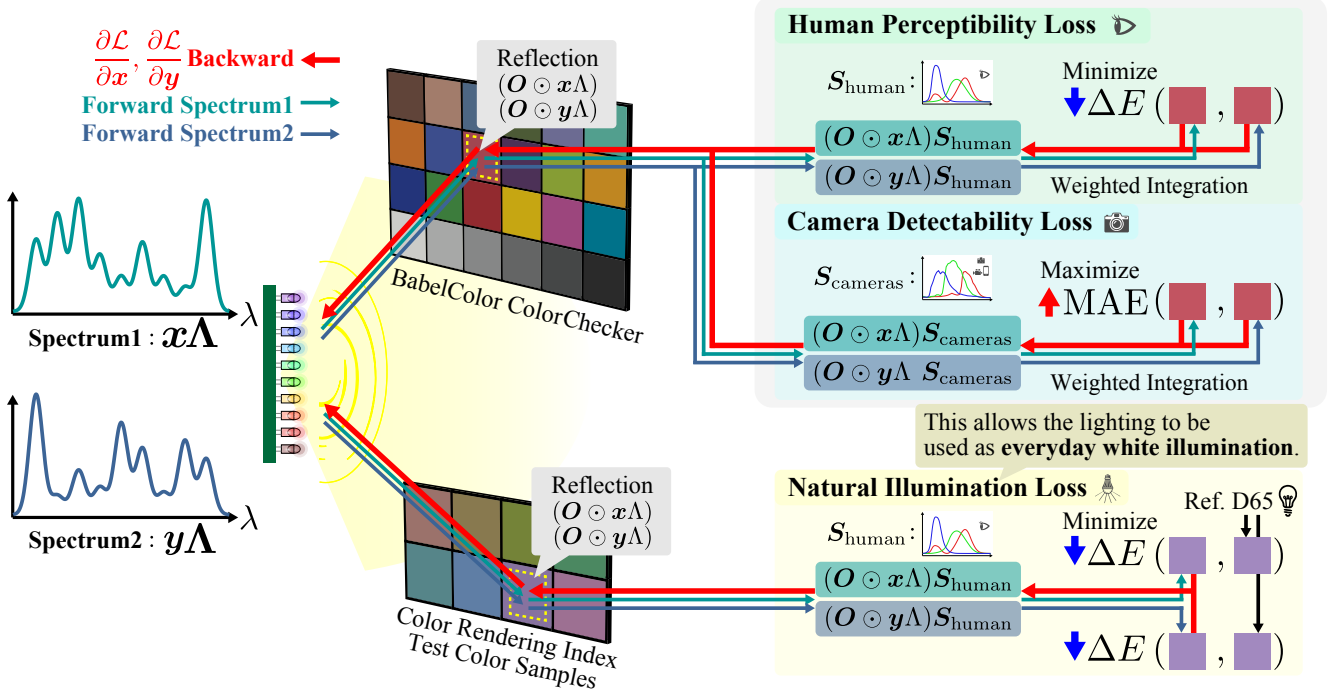


Figure 3. Proposed optimization pipeline. The objective functions ensure human imperceptibility, camera detectability, and natural white light for everyday use. These functions guide the optimization of LED intensity variables to achieve the desired lighting characteristics. Dark blue and dark green arrows indicate the forward pass for the two illuminations, while red arrows denote the backpropagation path used to optimize the LED intensities.

Comparing Camera and Human Spectral Sensitivity.

Consumer cameras approximate human color perception using three primary color filters. However, achieving an exact match with human visual sensitivity is challenging due to technical and design constraints. Factors such as filter material properties, manufacturing costs, and brand-specific design choices create variations in camera spectral sensitivity. Due to differences in spectral sensitivities, cameras apply color space mappings to convert sensor-specific RGB values to a human-perceptual color space [9, 14, 17]. Linear transformations are commonly used to align the camera’s color space with the human XYZ space, but these adjustments are inherently limited. As a result, camera sensors still differ from human vision in their spectral response [13].

3. Proposed Method

As discussed in Sec. 2, notable sensitivity differences exist between human vision and cameras. Leveraging these differences, we propose a method that optimizes spectral lighting to embed information detectable by cameras while remaining imperceptible to humans. This section details the spectral optimization process, the flicker detection technique for information retrieval, and the impact of noise on detection accuracy.

3.1. Spectral Optimization

We optimize two ambient light conditions that are imperceptible to humans but detectable by a camera to achieve communication through lighting. We denote these optimized ambient lights as L_1 and L_2 . Switching between these lights at a rate equal to or lower than the frame rate establishes communication between the lighting system and the cameras. The two ambient light conditions are optimized to satisfy the following three requirements jointly:

1. The difference between the lights is imperceptible to humans.
2. The difference is detectable by the cameras.
3. The ambient light provides a natural color rendering suitable for everyday use.

Let the reflectance characteristics of the objects be denoted as $R_o \in \mathbb{R}^{N \times l}$, where N is the number of objects. The reflectance of the i -th object is denoted as $R_o[i] \in \mathbb{R}^l$. Fig. 3 illustrates the optimization pipeline incorporating these requirements.

Human Perceptibility. The perceptibility of color differences to humans has been extensively studied in color difference research and is quantified using DeltaE2000 [3, 24]. For our method, we use DeltaE2000 (Δ_E) to ensure that the reflected light from all objects remains imperceptible to hu-

mans. To meet this criterion, we define the following evaluation function:

$$\mathcal{L}_h = \Delta_E(s_h(\mathbf{R}_o \odot \mathbf{L}_1), s_h(\mathbf{R}_o \odot \mathbf{L}_2)), \quad (1)$$

where s_h denotes the spectral sensitivity of the human visual system, $(\mathbf{R}_o \odot \mathbf{L}_i)$ represents the element-wise product of the object reflectance \mathbf{R}_o and the light spectrum \mathbf{L}_i , and $s_h(\mathbf{R}_o \odot \mathbf{L}_i) \in \mathbf{R}^3$ represents the observed XYZ values.

A difference of $\Delta_E \leq 2$ is generally imperceptible to untrained observers when colors are viewed side by side [3, 24]. However, in practical applications, the human eye is more sensitive to low-frequency flicker, making it more likely to detect subtle differences when colors alternate over time, even if these differences are imperceptible in a static, side-by-side comparison. Therefore, the evaluation function is minimized under this constraint, provided that the requirements of the other loss functions are satisfied.

Camera Detectability. To ensure the watermark remains imperceptible, it is crucial that the flicker does not produce excessive pixel differences that could be easily noticeable to viewers. If the flicker in the captured video is too prominent, it may significantly degrade visual quality, making the video unsuitable for typical viewing. The following evaluation function is designed to ensure adequate detectability across all spectral channels while controlling pixel differences for multiple camera models:

$$\mathcal{L}_c = -\min_i f(\text{MAE}(v_1[i], v_2[i])) \quad (2)$$

$$v_1[i] = s_c[i](\mathbf{R}_o \odot \mathbf{L}_1) \quad (3)$$

$$v_2[i] = s_c[i](\mathbf{R}_o \odot \mathbf{L}_2) \quad (4)$$

$$f(x) = \begin{cases} x & x < \tau_c \\ \tau_c & \text{otherwise} \end{cases} \quad (5)$$

where $s_c[i]$ denotes the spectral sensitivity of the i -th channel for each camera, MAE is the mean absolute error, and τ_c is a threshold. The function $f(x)$ clips the loss to a maximum of τ_c .

To align this objective with other losses in the optimization, which are formulated as minimization tasks, a negative sign is applied to \mathcal{L}_c . This converts it from a maximization objective to a minimization one, ensuring consistency across all loss terms.

The spectral sensitivity differences between humans and cameras vary across channels, and minimizing the average error across all channels can result in disproportionately high or low errors in specific ones. This approach prioritizes the channel with minimum error to prevent the embedding from being affected by an object's color or reflectance characteristics.

This formulation ensures that the flicker effect in the camera remains within a manageable range, keeping the average difference across all channels near the threshold τ_c .

We set τ_c to $1/256$, representing the detectable value in the context of 8-bit video quantization. This threshold allows for reliable detection, as the average includes colors with low brightness and those predominantly influenced by specific color channels, meaning not all pixel differences are strictly limited to $1/256$. Additionally, the loss function is averaged across multiple cameras to ensure broader applicability rather than optimizing for a single device.

Natural Illumination. Another criterion for this method is that the light must maintain a neutral white appearance with minimal spectral skew, ensuring suitability for everyday use. Color rendering is conventionally evaluated using the Color Rendering Index (CRI), which measures color fidelity based on standardized samples. CRI values range from 0 to 100, with higher values indicating more natural color rendering. Common household LEDs and bulbs typically have CRI values around 60, which is generally sufficient for everyday applications.

Our approach targets the standard D65 illumination (i.e., spectra that appear as daylight) and enforces a CRI of at least 60 per the CRI standard (eight test samples). We define $\text{CRI} = 100 - 4.6 \Delta E$, where ΔE is the average color difference to D65. To ensure $\text{CRI} \geq 60$, we minimize a hinge penalty on ΔE (Eq. 6).

$$\mathcal{L}_w = \sum_{i=1}^8 \text{ReLU}(\Delta_{1,i} - \tau_w) + \text{ReLU}(\Delta_{2,i} - \tau_w) \quad (6)$$

$$\Delta_{1,i} = \Delta E(s_h(\mathbf{R}_o[i] \odot \mathbf{L}_{\text{ref}}), s_h(\mathbf{R}_o[i] \odot \mathbf{L}_1)) \quad (7)$$

$$\Delta_{2,i} = \Delta E(s_h(\mathbf{R}_o[i] \odot \mathbf{L}_{\text{ref}}), s_h(\mathbf{R}_o[i] \odot \mathbf{L}_2)), \quad (8)$$

where \mathbf{L}_{ref} is the D65 reference spectrum and $\tau_w = 40/4.6$ enforces $\text{CRI} \geq 60$.

Weighted Optimization. The optimal two spectra are designed by adjusting and minimizing these three losses as follows:

$$\mathbf{L}_1, \mathbf{L}_2 = \arg \min_{\mathbf{L}_1, \mathbf{L}_2} (w_h \mathcal{L}_h + w_c \mathcal{L}_c + w_w \mathcal{L}_w), \quad (9)$$

where w_h, w_c, w_w are the weights for each loss function, which determine the optimization direction. These weights essentially involve a trade-off, so their selection is critical. We set the weights such that the color rendering loss becomes zero and the camera loss is approximately $1/256$. The human visual loss is minimized as much as possible while satisfying these conditions.

Reparameterization for Hyperspectral Illumination. As shown in Fig. 2, hyperspectral illumination can be reproduced as a weighted sum of multiple LEDs. Therefore, rather than directly optimizing \mathbf{L}_1 and \mathbf{L}_2 , we optimize the intensity of each LED. Let each LED's spectral profile be denoted as Λ_i and their respective intensities be x_i and y_i .

The spectral compositions of the lighting follow:

$$L_1 = \sum_{i=1}^m x_i \Lambda_i, \quad L_2 = \sum_{i=1}^m y_i \Lambda_i, \quad (10)$$

where m is the number of LEDs. In this manner, reparameterization shifts the optimization target to $\mathbf{x} = [x_0, \dots, x_m]$ and $\mathbf{y} = [y_0, \dots, y_m]$.

3.2. Detection of Flicker

We now outline the detection algorithm. The supplementary material provides a more detailed explanation.

Frames per Second (fps) and Bits per Second (bps). In our setup, lighting fluctuations are encoded as binary values, with frame-to-frame pixel intensity changes representing bits (1 for a change, 0 for no change). This differential encoding generates a bit sequence based on frame differences across the image. According to Shannon’s theorem, a recording rate of n fps ideally enables communication at n bps. However, real cameras have finite exposure times, which can introduce inaccuracies if phase shifts occur between illumination changes and the camera’s sampling phase. To mitigate this, we set the recording rate to $2n$ fps, ensuring synchronization and achieving n bps.

Signal in Dark Regions. In dark regions, 8-bit quantization can erase inter-frame differences, as pixel fluctuations are often too small to detect accurately. To improve detectability, we set the camera detectability loss so that the average pixel fluctuation is approximately 1. Pixels with values consistently below 128 across all channels and time are masked to prevent quantization loss.

Normalization. To ensure detection accuracy despite differences in scene brightness due to camera settings like exposure time and white balance, we normalize pixel intensity values across all channels, setting the mean intensity to 1. This normalization standardizes variations, enabling consistent results across different camera configurations.

Temporal Differential Calculation. With normalized data, we compute differences along the time axis to capture frame-to-frame changes representing encoded binary values. This is achieved by subtracting each frame from the previous one. At this stage, differential information is obtained at $2 \times$ sampling rate to mitigate phase shifts. For final processing, we return to the signal’s original sampling rate by downsampling, selecting the phase that best captures the lighting fluctuations to address potential misalignment.

Signal Binarization. Finally, we compute the average brightness changes across the image for each time frame to capture fluctuations over time. To extract a binary signal that adapts to the scene’s characteristics, we use an adaptive thresholding approach. We begin with a static threshold to

estimate the average brightness fluctuation level when significant changes occur. This serves as a baseline for assessing typical fluctuation intensity. Based on this baseline, we then define a dynamic threshold set to half of the average fluctuation, ensuring consistent identification of significant brightness changes as binary signals.

3.3. Noise Effects

The effect of noise during image capture is an important factor in determining the ability to detect flicker. As this algorithm averages the pixel differences across the entire image, the effect of noise is minimized. Assuming the noise is additive white Gaussian noise with a mean of 0 and variance σ^2 , the mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ of the noise relative to the overall image difference can be expressed as follows:

$$\tilde{\mu} = 0, \quad \tilde{\sigma} = \sigma \sqrt{\frac{2}{N}} \quad (11)$$

Considering that recent video resolutions are around 1 megapixel, the effect of noise is sufficiently negligible.

4. Experimental Results

This section describes the experimental setup and evaluations used to validate the proposed method. First, we detail the implementation of spectral optimization, ensuring embedded information is detectable by cameras while remaining imperceptible to human observers. Next, we assess human sensitivity to the optimized flicker to confirm its imperceptibility. Finally, we quantitatively evaluate the system’s communication accuracy across various environmental conditions and camera setups, examining its robustness and reliability in practical scenarios.

4.1. Spectral Optimization

We use a BabelColor ColorChecker chart with known reflectance properties to optimize the spectra. This chart consists of 24 color patches, each with a defined reflectance value. Spectral optimization is performed by treating the chart as the target object.

For the prototype lighting setup, we use ten types of LEDs with different wavelengths, as shown in Fig. 4. LED brightness is modulated by adjusting the current through a constant-current chip using pulse-width modulation via a Raspberry Pi. During optimization, the duty cycle of each LED serves as the adjustable variable for brightness control.

Camera sensor characteristics for the optimization are taken from a publicly available dataset of spectral sensitivities for 28 cameras [13]. The optimization problem is solved using the Adam optimizer [18] with a learning rate of 0.01, over 5000 iterations. The hyperparameters for optimization, denoted as $\mathbf{w}_h, \mathbf{w}_c, \mathbf{w}_w$, are determined through grid search. The hyperparameters used in this case are $\mathbf{w}_h = 0.15, \mathbf{w}_c = 0.05, \mathbf{w}_w = 0.8$.

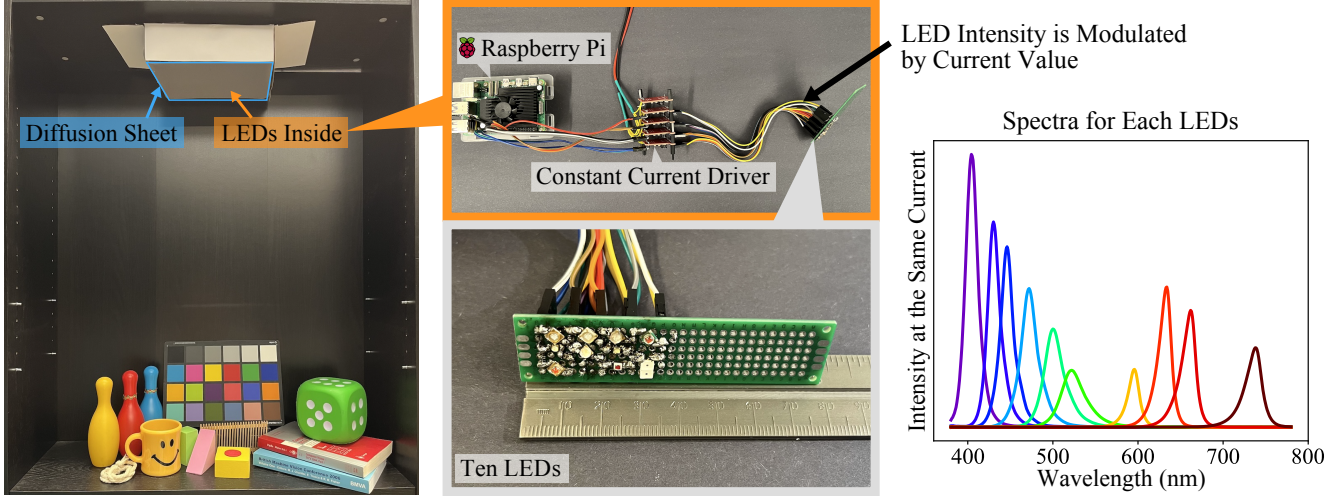


Figure 4. Lightbox using the prototype hyperspectral lighting. Each LED’s intensity is adjusted using a constant-current chip and a Raspberry Pi. The spectral distribution of each LED is measured inside the light box by capturing the white patch of the color chart.

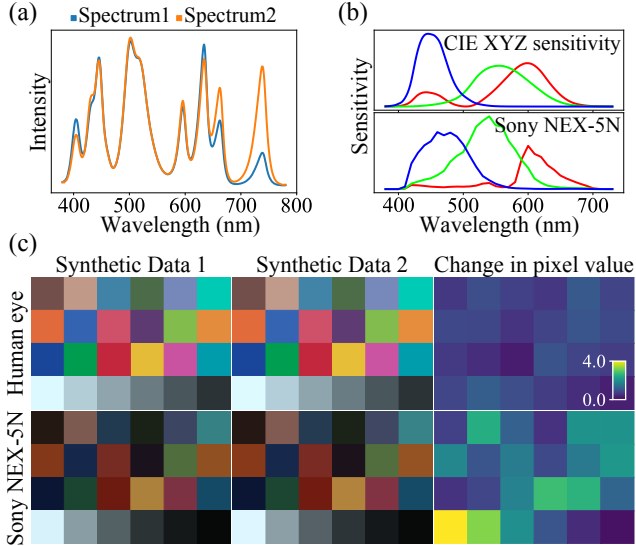


Figure 5. (a) Plot of the two optimized spectra. (b) The CIE 1931 Standard Observer XYZ Color Matching Functions and an example from the camera spectral sensitivity dataset (Sony NEX-5N). (c) Synthetic rendered color checker and pixel value variations obtained using the optimized spectra.

The final spectra obtained through this optimization process are shown in Fig. 5. The optimized pair yields a mean $\Delta E = 0.2839$, i.e., imperceptible.

Based on the cameras’ spectral sensitivities from the dataset, the average MAE across all channels was 1.843 in 8-bit representation. Using the proposed method’s definition of \mathcal{L}_c , the minimum MAE across channels is 0.972, sufficient to produce detectable pixel value differences even after 8-bit quantization.

4.2. Human Sensitivity Evaluation

Experimental Setup. To assess the perceptibility of the proposed method, we evaluated human color perception under LED illumination optimized using the proposed spectra. For this evaluation, we used the lightbox setup shown in Fig. 4, which contained various colored objects for participants to observe.

We prepared two flickering illumination patterns:

1. **Optimized Spectra:** Alternating illumination between the two LED-based spectra optimized by our method.
2. **Baseline Intensity Pattern:** A conventional VLC scheme that varies light intensity for data transmission (ratio 1:0.97).

To equate the sensor-level signal, the baseline ratio was tuned such that, under the same camera and exposure, the mean per-frame pixel-value difference between its two intensity levels matched that observed between the optimized spectra.

Participants took part in the experiment under the following sequence of conditions:

1. **Viewing Steady-State Baseline:** Participants first observed the objects under a steady optimized light for 10 sec to understand the arrangement of objects.
2. **Randomized Observation of Flicker Patterns:** Participants then observed a randomly selected lighting pattern from four possible conditions:
 - Optimized Steady State
 - Optimized Flicker State
 - Baseline Steady White Light
 - Baseline Flicker White Light (intensity ratio of 1:0.97)
3. **Feedback on Perceived Flicker:** Participants were asked whether they noticed any flicker in the lighting after each observation.



Figure 6. Our experimental setup shows a participant observing the optimized lighting patterns. Shielding prevents direct light exposure, allowing participants to observe the flicker indirectly through reflected light only.

4. Repeated Randomized Testing: This process was repeated so that each lighting pattern was presented exactly five times to each participant, ensuring a balanced distribution across all conditions.

Illumination Type	Avg. detections (range)	Detection rate
Optimized Steady	0.133 (0–2)	2.67%
Optimized Flicker	0.467 (0–2)	9.33%
Baseline Steady	0.267 (0–2)	5.33%
Baseline Flicker	4.933 (4–5)	98.67%

Table 1. Detection results. For each condition, 15 participants each performed five trials. “Avg. detections” is the average number of detections per participant over five trials (parentheses show the participant-wise min–max). “Detection rate (%)” equals $(\text{avg. detections} / 5) \times 100$.

Results. We tested 15 participants (Table 1). Each participant was exposed to four illumination conditions: (i) Optimized Steady State; (ii) Optimized Flicker State ($\Delta E = 0.2839$); (iii) Baseline Steady White Light; and (iv) Baseline Flicker White Light with an intensity ratio of 1:0.97 ($\Delta E = 0.6080$). Although ΔE values in this range are typically regarded as below the threshold of noticeability for static color differences, in our low-frequency flicker experiment, the Baseline Flicker White Light condition (1:0.97) was detected by nearly all participants (98.67%), indicating that such small chromatic differences become clearly visible when presented as low-frequency temporal flicker.

By contrast, the optimized spectral modulation yielded a mean detection rate of only 9.33%; moreover, no participant reported it in more than half of the five trials, indicating that it was effectively imperceptible for most observers.

For the static controls, false-positive rates were lower

still (2.67% for static optimized and 5.33% for static white), likely reflecting minor fluctuations or momentary adaptation. Taken together, these results support spectral optimization as a viable strategy for low-frequency modulation in imperceptible watermarking applications. At the same time, the findings underscore that conventional intensity modulation is unsuitable in this regime due to human flicker sensitivity.

4.3. Evaluation of Decoding

Communication Accuracy in a Single Scene. First, we evaluated the communication accuracy of the proposed method in a controlled environment using a single scene. We conducted long-duration recordings with three cameras (iPhone 13, Pixel 8 Pro, and Nikon D7200), each receiving 6400 bits, over several minutes per device. This setup established a baseline measure of communication accuracy for each camera and assessed whether the proposed method could achieve consistent performance across different devices. The results are shown in Fig. 7.

In this experiment, no bit errors were observed in the 6400-bit transmissions for any of the cameras, indicating that averaging across all pixels helps mitigate noise effects and ensures stable communication.

Additionally, zoomed-in areas reveal that while compression and differential representation vary between cameras, all successfully recovered the signal. Notably, on the Pixel 8, not all pixels exhibited positive fluctuations, yet the overall average variation across the frame remained positive, enabling correct signal recovery.

Method Robustness in Diverse Environments. Since practical applications involve diverse lighting conditions and scene compositions, additional tests were conducted to evaluate the proposed method’s robustness.

First, we used two smartphones to capture 128 bits of data in each of ten scenes to assess accuracy across different scenes. Examples of these scenes are shown in Fig. 8. This experiment evaluated the impact of scene diversity on detection accuracy and explored the range of environments where the proposed method could be effectively applied.

Additionally, to test robustness against variations in lighting conditions and camera angles, we adjusted the distance between the light source and the object to 88.9, 76.2, and 63.5 cm (=35, 30, and 25 in), positioning the smartphones at ten viewing angles for each distance. In each configuration, 128 bits of data were transmitted, allowing us to assess the impact of lighting and camera angles on communication accuracy.

No bit errors were observed under any tested conditions, indicating stable communication performance across diverse scenes and lighting setups.

Using dynamic thresholding, the proposed method demonstrated reliable signal recovery even in scenes with

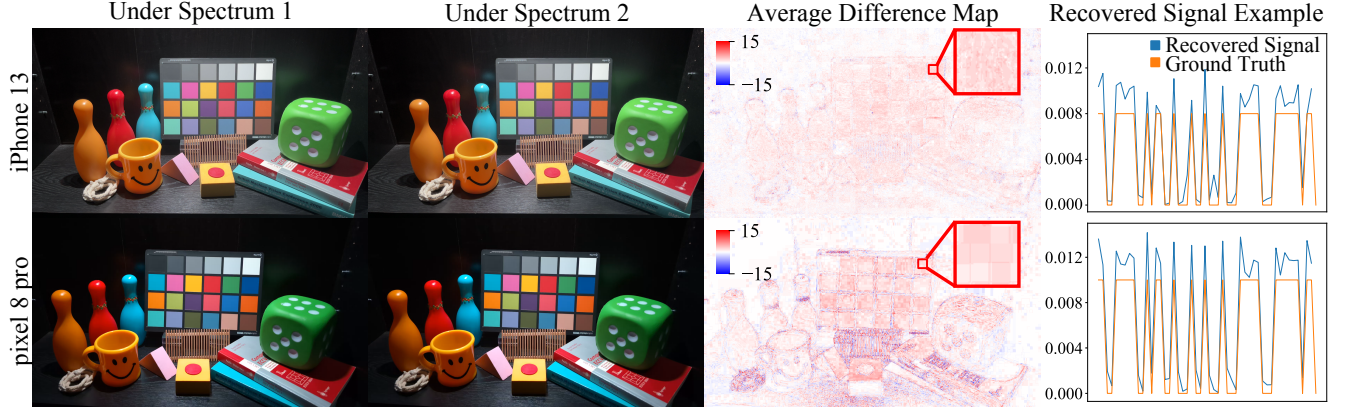


Figure 7. This figure illustrates the decoding results from a scene captured using two spectral lighting conditions, Spectrum 1 and Spectrum 2. The first two images display frames from the observed video—(left) Spectrum 1 and (right) Spectrum 2. The third image displays the average RGB differences between these frames. The fourth plot shows the recovered signal values before binarization.

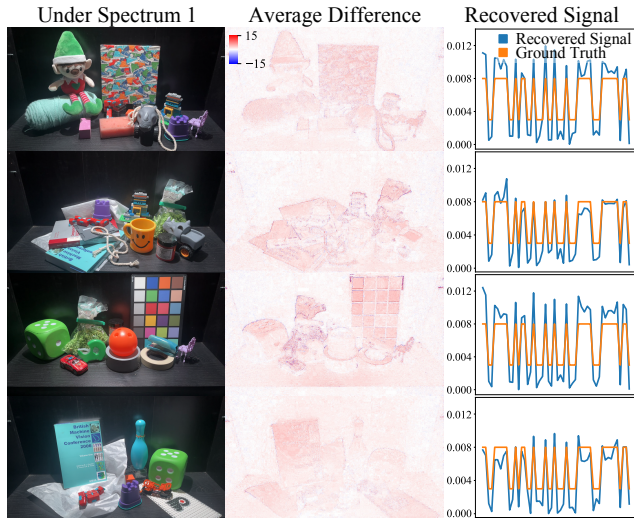


Figure 8. Experimental scenes were captured with iPhone 13 and evaluated, with objects of varying reflectance properties arranged as randomly as possible to assess robustness across lighting and reflectance conditions. The graphs in the figure display the recovered signal values before binarization.

varying reflectance distributions and viewing angles. As shown in Fig. 8, the recovered signal scales differently across scenes due to reflectance variations, even after normalizing pixel values. Dynamic thresholding ensures consistent signal recovery despite these scale differences, enabling accurate communication in diverse conditions.

5. Conclusion

We have presented the first method to leverage spectral modulation from narrowband LEDs for metadata embedding in video content—imperceptible to human vision yet

detectable by RGB cameras. Our approach optimizes spectral modulation to emulate D65 illumination while accounting for both camera and human spectral sensitivities. This enables real-time metadata embedding using ambient lighting without requiring camera modifications. Verification experiments confirmed that these modulations remain undetectable to humans and are robust across different cameras and environmental conditions.

Our findings suggest that spectral modulation from LED lighting holds significant potential for privacy protection and verification applications. By leveraging imperceptible spectral variations rather than intensity differences, our method enables non-intrusive, optically embedded watermarks that preserve the visual experience for human viewers while ensuring reliable detection by camera sensors.

Limitations and Future Directions. While our analysis focuses on static objects, the same conclusion typically holds under small motions when the global image sum is stable, since $\sum(I_{t-1} - I_t) = \sum I_{t+1} - \sum I_t$. As shown in the supplementary material, deviations arise when $\sum I_t$ changes—e.g., due to non-Lambertian/specular effects, ambient-light shifts, exposure/white-balance adjustments, or occlusions/parallax—which we leave for future evaluation and algorithmic refinements.

Another limitation of our approach is its relatively low communication speed, capped at 15 bps, which may limit data transfer in time-sensitive applications. One potential solution is to increase the data rate by parallelizing communication across multiple wavelength channels, enabling independent data transmission on each channel.

Beyond metadata embedding, imperceptible spectral modulation may have broader implications in optical sensing. While its potential is yet to be fully explored, it could contribute to areas such as hyperspectral imaging and 3D sensing, where spectral control plays a crucial role.

Acknowledgments

This work was partially supported by JST ASPIRE Grant No. JPMJAP2502 and JSPS KAKENHI Grant No. 24KJ1660.

References

- [1] George A. Agoston. *Color Theory and Its Application in Art and Design*. Springer, 2013. 2
- [2] Kiarash Amiri, Shih-Hsien Yang, Aditi Majumder, Fadi Kurdahi, and Magda El Zarki. Mobile collaborative video. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(9):1594–1604, 2014. 2
- [3] David H. Brainard. Color appearance and color difference specification. In *The Science of Color*, pages 191–216. Elsevier, 2003. 3, 4
- [4] Tu Bui, Shruti Agarwal, and John Collomosse. Trustmark: Robust watermarking and watermark removal for arbitrary resolution images. In *Int. Conf. Comput. Vis.*, pages 18629–18639, 2025. 2
- [5] Chi-Wai Chow, Chung-Yen Chen, and Shih-Hao Chen. Visible light communication using mobile-phone camera with data rate higher than frame rate. *Opt. Express*, 23(20):26080–26085, 2015. 2
- [6] *Proceedings of the Eighth Session of the Commission Internationale de l’Eclairage*, 1932. Commission Internationale de l’Eclairage, Cambridge University Press. 2
- [7] Mark D. Fairchild. *Color Appearance Models*. John Wiley & Sons, 2013. 2
- [8] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, 2011. 2
- [9] Guowei Hong, M. Ronnier Luo, and Peter A. Rhodes. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Res. Appl.*, 26(1):76–84, 2001. 3
- [10] Berthold K. P. Horn. Exact reproduction of colored images. *Comput. Vis. Graph. Image Process.*, 26:135–167, 1984. 2
- [11] IEEE. IEEE standard for local and metropolitan area networks—part 15.7: Short-range wireless optical communication using visible light. *IEEE Std 802.15.7*, pages 1–309, 2011. 2
- [12] Peng Ji, Hsin-Mu Tsai, Chao Wang, and Fuqiang Liu. Vehicular Visible Light Communications with LED Taillight and Rolling-Shutter Camera. In *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, pages 1–6, 2014. 2
- [13] Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *WACV*, 2013. 1, 3, 5
- [14] Hakki Can Karaimer and Michael S. Brown. Improving color reproduction accuracy on cameras. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018. 3
- [15] Dilukshan Karunatilaka, Fahad Zafar, Vineetha Kalavally, and Rajendran Parthiban. LED based indoor visible light communications: State of the art. *IEEE Communications Surveys & Tutorials*, 17(3):1649–1678, 2015. 2
- [16] Mohammad Ali Khalighi and Murat Uysal. Survey on free space optical communication: A communication theory perspective. *IEEE Commun. Surveys Tuts.*, 16(4):2231–2258, 2014. 2
- [17] Dongyoung Kim, Mahmoud Afifi, Dongyun Kim, Michael S. Brown, and Seon Joo Kim. Ccmnet: Leveraging calibrated color correction matrices for cross-camera color constancy. In *Int. Conf. Comput. Vis.*, 2025. 3
- [18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [19] Kevin Liang, Chi-Wai Chow, and Yang Liu. RGB visible light communication using mobile-phone camera and multi-input multi-output. *Opt. Express*, 24(9):9383–9388, 2016. 2
- [20] Jong-Il Park, Moon-Hyun Lee, Michael D Grossberg, and Shree K Nayar. Multispectral imaging using multiplexed illumination. In *Int. Conf. Comput. Vis.*, pages 1–8, 2007. 2
- [21] Abdullah Sevincer, Aashish Bhattarai, Mehmet Bilgi, Murat Yuksel, and Nezih Pala. Lightnets: Smart lighting and mobile optical wireless networks — a survey. *IEEE Commun. Surveys Tuts.*, 15(4):1620–1641, 2013. 2
- [22] Suhyun Shin, Seokjun Choi, Felix Heide, and Seung-Hwan Baek. Dispersed structured light for hyperspectral 3d imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 24997–25006, 2024. 2
- [23] Suhyun Shin, Seungwoo Yoon, Ryota Maeda, and Seung-Hwan Baek. Dense dispersed structured light for hyperspectral 3d imaging of dynamic scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16589–16598, 2025. 2
- [24] Mike Stokes, Mark D. Fairchild, and Roy S. Berns. Precision requirements for digital color reproduction. *ACM Trans. Graph.*, 11(4):406–422, 1992. 3, 4
- [25] Ferenc Szabó, Janos Schanda, Peter Bodrogi, and Emil Radkov. A comparative study of new solid state light sources. In *CIE Session*, 2007. 2
- [26] Yuichi Tanaka, Shinichiro Haruyama, and Masao Nakagawa. Wireless optical transmissions with white colored led for wireless home links. In *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1325–1329, 2000. 2
- [27] Dobrosław Tsonev, Stefan Videv, and Harald Haas. Light fidelity (Li-Fi): towards all-optical networking. In *Proc. SPIE OPTO*, page 900702, 2014. 2
- [28] Guenther Wyszecki and Walter Stanley Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulas*. Wiley, 1982. 2
- [29] Zhihao Xia, Jason Lawrence, and Supreeth Achar. A dark flash normal camera. In *Int. Conf. Comput. Vis.*, pages 2430–2439, 2021. 2
- [30] Xuanyu Zhang, Zecheng Tang, Zhipei Xu, Runyi Li, Youmin Xu, Bin Chen, Feng Gao, and Jian Zhang. Omniguard: Hybrid manipulation localization via augmented versatile deep image watermarking. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3008–3018, 2025. 2
- [31] Yinda Zhang, Sameh Khamis, Christoph Rhemann, Julien Valentin, Adarsh Kowdle, Vladimir Tankovich, Michael

Schoenberg, Shahram Izadi, Thomas Funkhouser, and Sean Fanello. Activestereonet: End-to-end self-supervised learning for active stereo systems. In *Eur. Conf. Comput. Vis.*, pages 784–801, 2018. [2](#)

- [32] Ming Zhao, Qinyuan Zhang, and Zhiguo Xia. Narrow-band emitters in led backlights for liquid-crystal displays. *Materials Today*, 40:246–265, 2020. [2](#)
- [33] Yinqiang Zheng, Ying Fu, Antony Lam, Imari Sato, and Yoichi Sato. Separating fluorescent and reflective components by using a single hyperspectral image. In *Int. Conf. Comput. Vis.*, 2015. [2](#)

Towards Imperceptible Watermarking Via Environment Illumination for Consumer Cameras

Supplementary Material

This supplementary material provides:

- A detailed explanation of the signal detection algorithm based on flicker modulation.
- Comprehensive data from human perceptibility evaluation.
- Details of the experiments conducted to validate detection accuracy.
- Dynamic Environments.

A. Detailed Implementation of Signal Detection Algorithm

This section provides a more detailed explanation of the signal detection algorithm that could not be fully described in the main paper.

Frames per Second (fps) and Bits per Second (bps) In our setup, lighting fluctuations are encoded as binary values, with each pixel value changing between consecutive frames representing bits (differential encoding). Specifically, if a pixel's intensity changes between frames, it is assigned a binary value of 1; if no change is detected, it is assigned a value of 0. This approach creates a bit sequence based on frame-to-frame differences across the image. According to Shannon's sampling theorem, a recording rate of n fps is theoretically required to achieve a communication speed of n bit/s. However, the theorem assumes instantaneous sampling, while real cameras observe over a finite exposure time. Therefore, if there is a half-period phase shift between the illumination change and the camera's sampling phase, the average of the fluctuations will always be observed, which may result in inaccurate measurements. To avoid the effects of phase shifts, we set the recording rate to $2n$ fps, achieving n bit/s while maintaining synchronization.

Signal in dark regions In differential encoding, frame-to-frame differences are calculated. In dark regions, these differences are often lost during quantization; thus, applying an appropriate mask enhances decoding accuracy. To improve detectability, we exclude pixels that are consistently dark in all channels over a set period, removing them from further processing. Consequently, the data is reduced from $\mathbf{V} \in \mathbf{R}^{T \times H \times W \times C}$ to $\mathbf{V}_m \in \mathbf{R}^{T \times N \times C}$, where N denotes the number of pixels that are not excluded.

Normalization To ensure detection accuracy despite differences in scene brightness due to camera settings like exposure time and white balance, we normalize pixel inten-

sity values across all channels, setting the mean intensity to 1. This normalization standardizes variations, enabling consistent results across different camera configurations, as follows:

$$\mathbf{V}_n[t, i, c] = \frac{\mathbf{V}_m[t, i, c]}{\frac{1}{TN} \sum_{j=0}^T \sum_{k=0}^N \mathbf{V}_m[j, k, c]} \quad (12)$$

where the numerator $\mathbf{V}_m[t, i, c]$ represents the pixel intensity at time t , position i , and channel c . The denominator $\frac{1}{TN} \sum_{j=0}^T \sum_{k=0}^N \mathbf{V}_m[j, k, c]$ calculates the mean pixel intensity for each channel.

Temporal differential calculation With normalized data, we next calculate the differences along the time axis to capture frame-to-frame changes that represent encoded binary values. This is done by subtracting each frame from the previous one:

$$\mathbf{D}[t, i, c] = \mathbf{V}_n[t, i, c] - \mathbf{V}_n[t-1, i, c] \quad (13)$$

At this stage, the computed differential information is obtained at a doubled sampling rate to reduce the impact of phase shifts. For the final processing, however, we return to the signal's original sampling rate. Specifically, to address potential phase misalignment, we downsample by selecting the phase that best captures the lighting fluctuations.

Signal extraction Finally, we calculate a measure of brightness changes across the entire image per time frame by averaging across all dimensions except time. Applying a threshold to the absolute values of these brightness fluctuations yields the signal. However, since this threshold depends on scene content, we first compute the average fluctuation value when changes occur using a static threshold τ_s . From this, a dynamic threshold τ_d is calculated as follows:

$$\tau_d = \frac{\mu_d}{2} \quad (14)$$

$$\mu_d = \frac{\sum_{\{t \mid \mathbf{D}'[t] > \tau_s\}} \mathbf{D}'[t]}{|\{t \mid \mathbf{D}'[t] > \tau_s\}|}, \quad (15)$$

where, $\mathbf{D}'[t]$ represents the downsampled difference value at time t , and $|\{t \mid \mathbf{D}'[t] > \tau_s\}|$ denotes the number of pixels exceeding τ_s .

B. Comprehensive Data from Human Perceptibility Evaluation

This section presents the complete data from the human perceptibility evaluation, including individual participant re-

sponses, to provide a detailed view beyond the summary statistics reported in the main text.

Detailed Participant Response Data Table 2 shows the detection counts for each participant under the tested lighting conditions. This table offers raw data, revealing individual variability and detection patterns across the following lighting conditions:

- **P1: Optimized Steady State** – A non-flickering lighting condition using optimized spectra.
- **P2: Optimized Flicker State** – Flickering lighting condition with optimized spectra.
- **P3: Baseline Steady White Light** – A non-flickering white light as a control.
- **P4: Baseline Flicker White Light** – Flickering white light with an intensity ratio of 1:0.97.

Participant ID	P 1	P 2	P 3	P 4
1	0	0	0	5
2	0	1	1	5
3	0	0	0	5
4	0	0	2	4
5	0	1	0	5
6	0	0	0	5
7	0	2	0	5
8	0	0	0	5
9	0	0	0	5
10	0	0	0	5
11	0	1	0	5
12	0	0	0	5
13	2	2	1	5
14	0	0	0	5
15	0	0	0	5

Table 2. Detection counts for each participant across four lighting patterns.

Analysis of Detection Patterns The data reveals that Pattern 4 consistently registered high detection counts across all participants, indicating its visibility was significantly more noticeable compared to the other patterns. Patterns 1, 2, and 3, on the other hand, showed much lower detection rates, suggesting that these patterns were largely imperceptible to human observers under the conditions tested. Additionally, none of the participants were able to detect all instances of flicker for Patterns 1, 2, or 3, reinforcing the conclusion that these patterns were effectively imperceptible.

C. Details for Evaluation of Decoding

This section provides the details of the experimental setup used for the quantitative evaluation that could not be fully

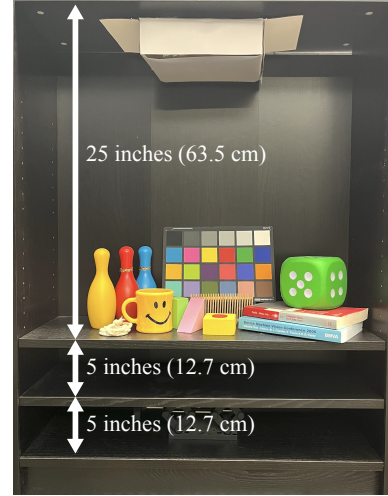


Figure 9. Setup for the experiment conducted at different heights.

described in the main paper. Additionally, it includes examples of the actual scenes captured during the experiments. The supplementary zip file also includes videos of the actual scenes recorded, offering further insights into the experimental setup and results.

Experimental Setup for Quantitative Evaluation Fig. 9 shows the experimental setup used to evaluate the method’s robustness under varying lighting conditions and camera angles. The experiment tested how changes in illumination angles at three different heights affected detection accuracy.

Results from Scene Variations Fig. 10 presents a selection of 10 different scenes captured to evaluate the method’s performance under varying scene conditions. The results include difference maps and recovered signals for each scene. As discussed in the main paper, while the absolute signal values vary across scenes, and some pixels show negative fluctuations, the proposed method successfully detects the signal in all cases.

C.1. Dynamic Environments (Simple Analysis)

Our decision aggregates frame-wise image differences over the whole image. Here we give an intuitive explanation *without* thresholds or absolute values. Let Ω be the pixel set and define the image sum

$$S_t = \sum_{p \in \Omega} I_t(p).$$

Then the total change due to object/scene variation between two frames is

$$\sum_{p \in \Omega} I_{t+1}(p) - I_t(p) = \sum_{p \in \Omega} I_{t+1}(p) - \sum_{p \in \Omega} I_t(p) \quad (16)$$

$$= S_{t+1} - S_t. \quad (17)$$

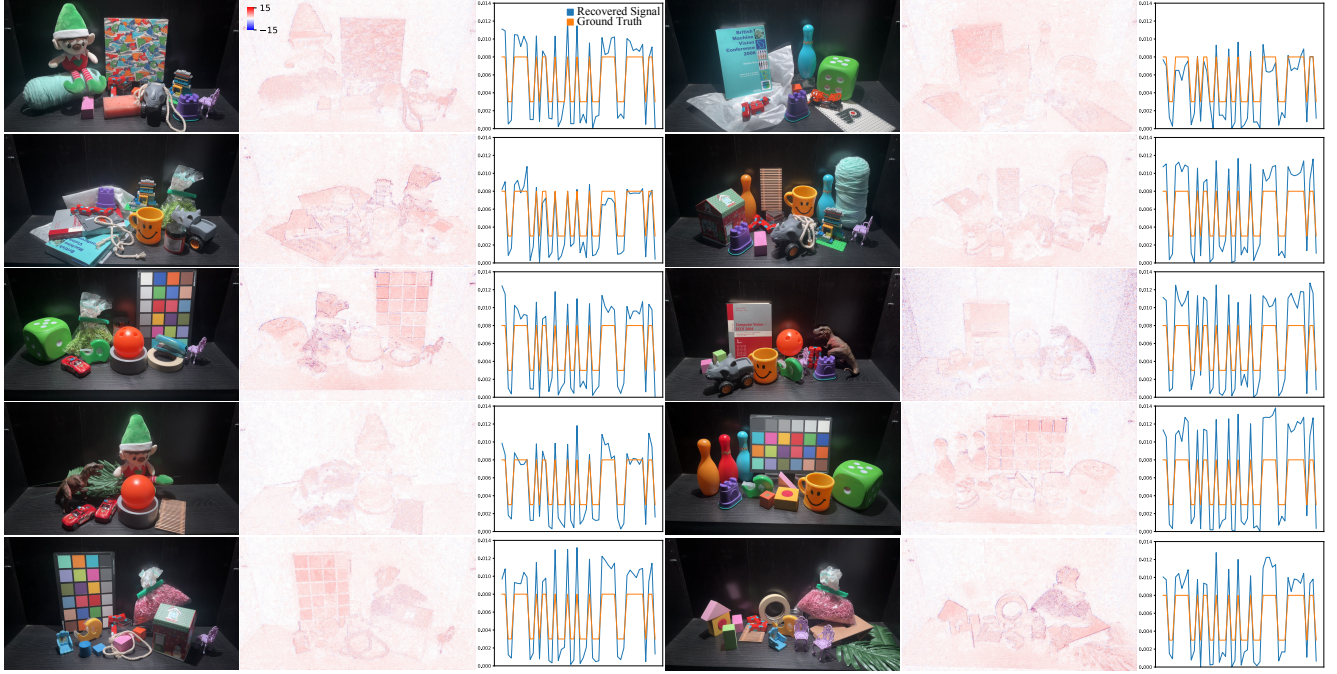


Figure 10. Captured results for 10 scenes used in the experiment, including difference maps and recovered signals.

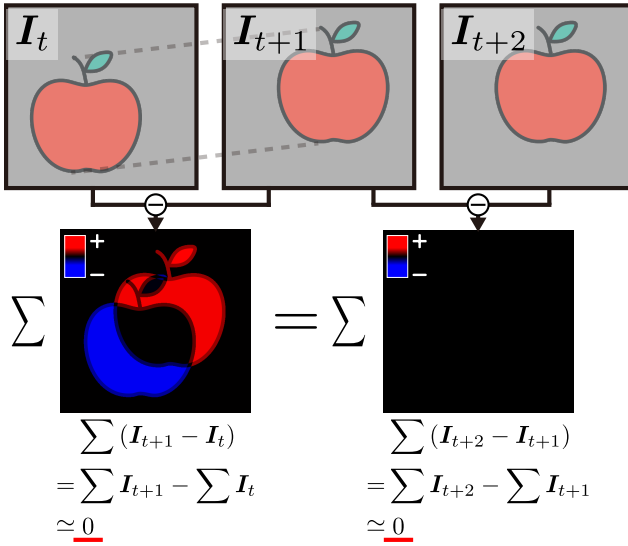


Figure 11. Variations caused by object shifts are canceled.

Therefore, as long as the global sum S_t does not vary much between frames (e.g., fixed exposure and approximately intensity-preserving conditions), we have $S_{t+1} - S_t \approx 0$ and the outcome is **essentially the same as for static objects**.

Supplementary Fig. 11 visualizes this: even with small object/scene motions, S_t remains nearly constant and thus $S_{t+1} - S_t$ stays small. Conversely, view-dependent specu-

larities, ambient-light changes, or exposure adjustments can alter S_t and lead to deviations from this simple behavior.