# Spatially Varying Radiometric Calibration for Camera-Display Messaging

Wenjia Yuan, Kristin J. Dana
Department of Electrical and Computer Engineering
Rutgers University
{wenjiay@eden, kdana@ece}.rutgers.edu

Ashwin Ashok, Marco Gruteser, Narayan Mandayam
WINLAB (Wireless Information Network Laboratory)
Rutgers University
{aashok,gruteser,narayan}@winlab.rutgers.edu

***Abstract*—Modern society has ubiquitous electronic displays including billboards, signage and kiosks. The concurrent prevalence of handheld cameras creates a novel opportunity to use cameras and displays as communication channels. The electronic display in this channel serves a twofold purpose: to display an image to humans while simultaneously transmitting hidden bits for decoding by a camera. Unlike standard digital watermarking, the message recovery in camera-display systems requires physics-based modeling of image formation in order to optically communicate hidden messages in real world scenes. By modeling the photometry of the system using a camera-display transfer function (CDTF), we show that CDTF depends on camera pose and varies spatially over the display. We devise a radiometric calibration to handle the nonlinearities of both the display and the camera, and we use the method for recovering video messages hidden within display images. Results are for 9 different display-camera systems for messages with 4500 bits. Message accuracy improves significantly with calibration and we achieve accuracy near 99% in our experiments, independent of the type of camera or display used.**

***Index Terms*—photometric modeling, radiometric calibration, spatially variations, convex optimization**

## I. Introduction

The recent prevalence of cameras and electronic displays provides the opportunity for a new type of communication channel called *visual MIMO* where the display pixels are transmitters and the camera pixels are receivers [1]–[3]. In this paper, we develop a method for sending and retrieving hidden time-varying messages using electronic displays and cameras that accounts for the physical model of image formation in a camera-display system. We assume the electronic display has two simultaneous purposes: 1) the original display function such as advertising, maps, or artwork; 2) the transmission of hidden time-varying messages.

When light is emitted from a display, the resultant 3D light field has an intensity that depends on the angle of observation as well as the pixel value controlled at the display. The emittance function of the electronic display is analogous to the BRDF of a surface because it characterizes the light radiating from a display pixel as a function of viewing angle. Messaging with displays is challenging because the emittance function varies depends on viewing angle and it varies spatially over the electronic display surface. Additionally, the emittance function has a particular spectral shape that does not match the spectral sensitivity curve of the camera. We combine the effects of the display emittance function and the spectral sensitivity of the camera into one system transfer function, as a *camera-display transfer function* (CDTF), which determines the captured pixel value as a function of the display pixel value. Our photometric model for image formation is shown in Figure 1. By using frame-to-frame characterization of the CDTF, the method is independent of the particular choice of display and camera.

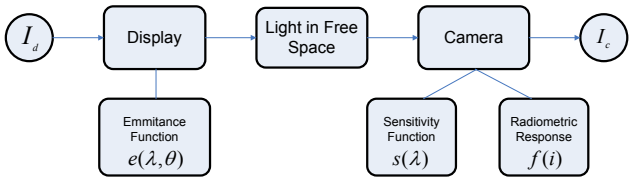


Fig. 1. Image Formation Pipeline: The image $\mathbf{I_d}$ is displayed by an electronic display with an emittance function $e$. The display is observed by a camera with sensitivity $s$ and radiometric response function $f$ to obtain the captured image $I_c$.

For CDTF estimation, we propose using textured patches placed within the display image that have intensity variation over the full range of display brightness values. We use the term *ratex patch* to refer to the radiometric calibration texture patches. These patches can be placed in corners as shown in Fig. 2. However, since these patches need only be an area of uniformly distributed intensity, nearly invisible patches can be created by using histogram equalization on corners of the original display image. The ratex patches have the advantage that they are perceptually acceptable, they represent the entire range of gray-scale intensity variation, and they can be distributed spatially. They are used for updating the spatially varying radiometric response function for each video frame, a necessary property since the CDTF depends on viewing angle and changes as the camera moves.

Our experimental results show that accuracy levels for message recovery can approach near 100% using our calibration approach. An evaluation of results has been provided by using video messaging with 40,500 message bits over 9 different combinations of commercial cameras and displays. We also explore the use of ratio versus additive methods of message embedding and discuss the strengths and weaknesses of each approach.

(a) (b) (c)

Fig. 2. Ratex patches placed in corners and used for radiometric calibration and classification training. (a) Original Image. (b) Ratex patches placed directly in image. (c) Ratex patches constructed by histogram equalization of each corner region, followed by image blending.

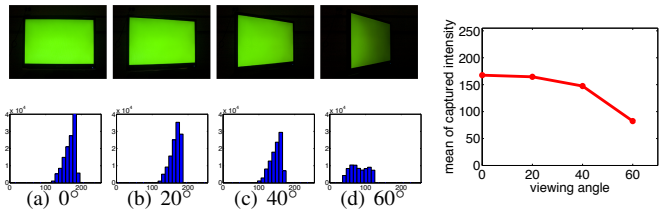


Fig. 3. An example of a displayed image with a single RGB color ($\rho_r = 0, \rho_g = 190, \rho_b = 30$), illustrating the dependence of the CDTF on viewing angle. **Top row, Left:** Captured images at four different viewing angles. **Bottom row, Left:** Intensity histograms of the captured images. **Right:** Mean values of the image intensity as a function of viewing angle.

## II. Related work

Since our work deals with hidden imagery, a related area is digital watermarking [4], [5]. Many watermarking techniques have been developed for robustness to geometric changes such as scaling, rotations, translations and general homography transformations [6]–[8]. However, the *photometry* of imaging has largely been ignored. The rare mention of photometric effects [9] [10] in the watermarking literature does not define photometry with respect to illumination; instead photometric effects are defined as "lossy compression, denoising, noise addition and lowpass filtering". In fact, photometric attacks are sometimes defined as jpeg compression [7].

Another related area is radiometric calibration, which estimates the camera response function that converts irradiance to pixel values. Many methods [11], [12] use multiple exposure times since light intensity on the sensor is a linear function of the exposure time. In our work, we are interested in the entire camera-display system that converts pixel values at the display to scene radiance and then converts scene radiance to camera pixels. We measure the overall response function (CDTF) using ratex patches present in each frame. Although more complex color models have been developed [13]–[15] for radiometric calibration, we have found the independent channel approach suitable for the display-camera representation.

Existing camera-display communications methods differ from our proposed approach. For example, invisible messages are applied in [16] and Bokode project [17], but these messages are fixed. LCD-camera communications is presented in [18] with visible time-varying messages, but the camera is in a fixed position with respect to the display. Recent work has been done in high speed visible light communications [19], but this work does not utilize existing displays and cameras and requires specialized hardware and LED devices. A novel method to communicate with cameras and displays exploits the rolling shutter of cameras to detect hues switched at 60Hz as described in [20]. Our method does not depend on the specific display characteristics, since such dependence are removed by using ratex patches for calibration in an online, frame-to-frame, manner.

## III. Methods

The captured image $\mathbf{I_c}$ from the camera viewing the electronic display image $\mathbf{I_d}$ can be modeled using the image formation pipeline in Fig. 1. When the display shows the value $(\rho_r, \rho_g, \rho_b)$ at a pixel in $\mathbf{I_d}$, it is emitting light in a manner governed by its emittance function and modulated by $\rho$. The emittance function $\mathbf{e} = (e_r, e_g, e_b)$ is typically a function of the viewing angle $\theta = (\theta_v, \phi_v)$ comprised of a polar and azimuthal component. For example, the emittance function of an LCD monitor has a large decrease in intensity with polar angle (see Fig. 3). Therefore the emitted light $I$ as a function of wavelength $\lambda$ for a given pixel $(x, y)$ on the electronic display is given by

$$I(x, y, \lambda) = \rho \cdot \mathbf{e}(\lambda, \theta). \tag{1}$$

The captured image $\mathbf{I_c}$ at the camera has three color components $(I_c^r, I_c^g, I_c^b)$. Now consider the intensity of the light received by one pixel element at the camera sensor. Let $s(\lambda) = (s_r, s_g, s_b)$ denote the camera sensitivity function for red, green and blue components. The captured image $\mathbf{I_c}$ can be represented as

$$\mathbf{I_c} \propto \int_\lambda [\rho \cdot \mathbf{e}(\lambda, \theta)] \cdot \mathbf{s}(\lambda) d\lambda. \tag{2}$$

The pixel value $\rho$ is controllable at the display, so we modify the display intensity by adding/multiplying the value $\kappa$ and transmit two consecutive images, one with the modified value $\mathbf{I_e}$ and one image of original intensity $\mathbf{I_o}$.

For the **additive-based** method, the embedded message is done by adding $\kappa$ as follows:

$$\mathbf{I_e} \propto \int_\lambda [(\kappa + \rho) \cdot \mathbf{e}(\lambda, \theta)] \, \mathbf{s}(\lambda) d\lambda. \tag{3}$$

Recovery of the embedded signal leads to a difference equation, which depends on the display emittance function and camera sensitivity function:

$$\mathbf{I_e} - \mathbf{I_o} \propto \int_\lambda [(\kappa) \cdot \mathbf{e}(\lambda, \theta)] \, \mathbf{s}(\lambda) d\lambda. \tag{4}$$

For the **ratio-based** method, the embedded message is done by multiplying $\kappa$:

$$\mathbf{I_e} \propto \int_\lambda [(\kappa \times \rho) \cdot \mathbf{e}(\lambda, \theta)] \, \mathbf{s}(\lambda) d\lambda. \tag{5}$$

Recovery of the embedded signal leads to a ratio equation

$$\mathbf{I_e}/\mathbf{I_o} \propto (\kappa). \tag{6}$$

The dependence on the properties of the display $e$ and the spectral sensitivity of the camera $s$ is removed.

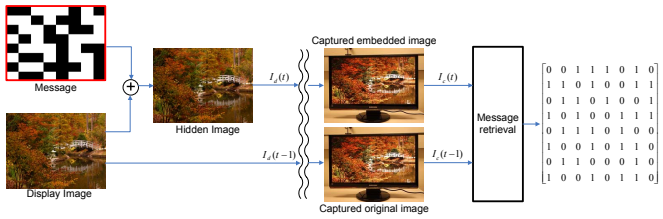


Fig. 4. Message Embedding and Retrieval. Two sequential frames are sent, an original frame and a frame with an embedded message image. Simple differencing is not sufficient for message retrieval. Our method is used to recover messages accurately.

The main concept for message embedding is illustrated in Fig. 4. In order to convey many "bits" per image, we divide the image region into a series of block components. Each block can convey a bit "1" or "0". For additive-based messaging, the blocks corresponding to a "1" contain the added value typically set to $\kappa = 5$ in an 8-bit image, while the zero blocks have no additive component ($\kappa = 0$). A small gray level is chosen to keep the messaging invisible to the human eye. The display can be tracked with existing methods [21]. For ratio-based messaging, the blocks corresponding to a "1" typically have $\kappa = 0.97$, while the zero blocks correspond to $\kappa = 1$.

When accounting for the nonlinearity in camera and display, we include the radiometric response function $\mathbf{I_c} = f(\mathbf{I_d})$, and the recovered display intensity is

$$\mathbf{I_d} = f^{-1}(\mathbf{I_c}) = g(\mathbf{I_c}). \tag{7}$$

We follow the approach of linear least squares [11] to represent the radiometric inverse function $g(i)$. The same inverse function $g$ is used for all color channels and gray-scale ratex patches are employed. This simplification of the color problem is justified by the accuracy of the empirical results.

To demonstrate the spatial variation present in typical systems, we find the CDTF for nine different display-camera combinations as shown in Fig. 5. We measure the CDTF by finding individual curves for each of 4 ratex patches. The color coding in Fig. 5 is the same as in Fig. 6. Notice the large spatial variation as indicated by the non-overlapping curves.

To handle spatial variance, our approach uses the ratex patches to find a calibration curve, or radiometric inverse function $g(i)$, for each corner. These curves are interpolated spatially in order to find a CDTF curve for any point on the captured display image.

## IV. Experiments

**Dataset and implementation details** For empirical validation, 9 different combination of displays and cameras are used, comprised of 3 displays: LG, Samsung SyncMaster, iMac; and 3 cameras: Canon EOS Rebel XSi, Nikon D70 Sony DSC-RX100. 15 display images are used, as illustrated in Fig. 7. From each display image, we create a display video of 10 frames: 5 frames with the original display images interleaved with 5 images of embedded time-varying messages. An embedded message frame is followed by an original image frame

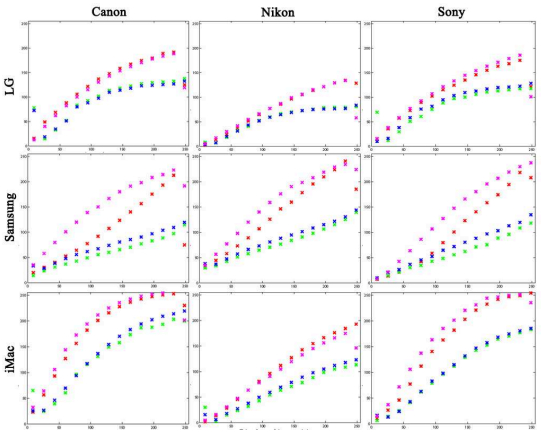


Fig. 5. Spatial variation of nine combinations of cameras and displays. The **x-axis** is the displayed intensities and **y-axis** represents the captured intensities. The curves show the CDTF. The four colors correspond to individual corners, which is the same color coding used in Fig. 6. Three displays are in rows and three cameras are in columns.

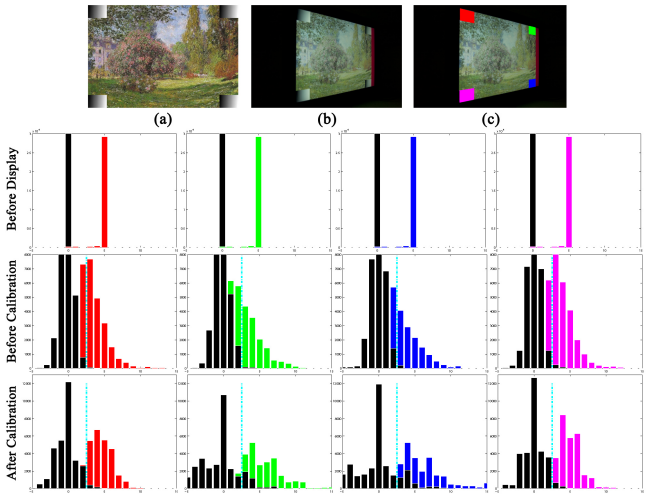


Fig. 6. Spatial distribution of ratex patches. **First Row: (a)** The displayed image. **(b)** The captured image with embedded bits. **(c)** Color coded ratex patches to identify the spatial location in the plots. (No colored ratex patches are displayed for the experiments; the color is for plot interpretation only). **Second row**: Histogram of the the ideal difference values, $I_e - I_o$ for label 0 (black) and label 1 (colors). **Third row**: The histogram of difference values at the ratex patch regions for label 0 (black) and label 1 (colors) from the captured image, before radiometric calibration. **Fourth row**: The histogram of difference values at the ratex patches regions for label 0 (black) and label 1 (colors) from the captured image after radiometric calibration. The blue dotted line shows the threshold value, which is reasonable only after calibration. Each patch is calibrated with a different curve, accounting for the intrinsic spatial variation.

to provide the temporal image pair $I_e$ and $I_o$. The display image does not change in the video, only the bits of the message frames. Each message frame has $8 \times 8 = 64$ blocks used for message bits (with 4 bits used for ratex patches for calibration and classification training data).

The accuracy for each video is defined as the number of correctly classified bits divided by the total bits embedded and is averaged over all testing videos. The entire test set over all display-camera combinations is 40,500 test bits. The experiments are conducted with a viewing angle of $45°$.

**Method 1** for message recovery uses no radiometric calibra-


Fig. 7. 15 display images for experiments. For each tested video sequence, the display image stays the same, but the hidden message changes over time.

tion. For additive embedding in Table I, the value $\kappa/2$, i.e. 2.5, is used as the threshold for classification. For ratio embedding in Table II, $\kappa/2$, i.e. 0.985, is used for classification. Method 1 can reach high accuracies in approximately linear systems, like iMac-Nikon as shown in Fig. 5. For this combination, accuracies for the additive and ratio embeddings are 98.56% and 99.04% respectively. A poor performance example is the system with the LG display. CDTF curves are non-linear and no accuracy over 90% is achieved for either method.

| | Canon | Nikon | Sony |
|---|---|---|---|
| Method 1 (LG Display) | 86.70 | 87.30 | 82.33 |
| Method 2 (LG Display) | **97.57** | **99.96** | **96.78** |
| Method 1 (Samsung Display) | 90.07 | 95.00 | 97.56 |
| Method 2 (Samsung Display) | **100** | **100** | **100** |
| Method 1 (iMac Display) | 94.00 | 98.56 | 97.63 |
| Method 2 (iMac Display) | **99.66** | **99.89** | **99.77** |

TABLE I
**ADDITIVE** EMBEDDING RESULTS (PERCENTAGE OF BITS CORRECTLY CLASSIFIED) USING METHOD 1(NO CALIBRATION) AND METHOD 2 (OUR PROPOSED CALIBRATION METHOD).

**Method 2** is our proposed method. Ratex patches are used to get four calibration curves at the corners, and the calibration curve for all non-ratex patch pixels is the linear combination of ratex patch curves. After calibration, the same thresholds as in Method 1 are applied for classification. The accuracy with our approach is higher than that of Method 1 in every case. The improvements are in the range of 0.25% to 16.93%.

Although ratio and additive methods have similar aggregate results, the ratio-based method performs poorly in dark regions of the display image because of quantization effects. That is, applying the ratio (typically 97%) to a low intensity value changes the image by at most 1-2 pixel levels which is not reliably detected.

## V. CONCLUSION

We have demonstrated that very high accuracy can be obtained using simple message embedding methods. However, the naive approach of thresholding the captured image is not sufficient. Radiometric calibration that varies spatially is the key to getting high accuracy results. Ratex patches in the display image provide useful calibration data. When training data can be used, support vector machine classifiers can be used to achieve similar accuracy. Consequently, near invisible messaging for display-camera communication can be achieved without the need for specialized hardware.

| | Canon | Nikon | Sony |
|---|---|---|---|
| Method 1 (LG Display) | 73.74 | 85.56 | 89.74 |
| Method 2 (LG Display) | **90.67** | **97.93** | **96.19** |
| Method 1 (Samsung Display) | 97.74 | 96.67 | 98.56 |
| Method 2 (Samsung Display) | **99.56** | **98.85** | **98.81** |
| Method 1 (iMac Display) | 86.89 | 99.04 | 96.37 |
| Method 2 (iMac Display) | **99.41** | **99.17** | **99.30** |

TABLE II
**RATIO** EMBEDDING RESULTS USING METHOD 1(NO CALIBRATION) AND METHOD 2 (OUR PROPOSED CALIBRATION METHOD).

## REFERENCES


[1] A. Ashok, M. Gruteser, N. Mandayam, J. Silva, M. Varga, and K. Dana, "Challenge: mobile optical networks through visual mimo," *MobiCom: Proceedings of the sixteenth annual international conference on Mobile computing and networking*, pp. 105–112, 2010.

[2] A. Ashok, M. Gruteser, N. Mandayam, and K. Dana, "Characterizing multiplexing and diversity in visual mimo," *Information Sciences and Systems (CISS)*, pp. 1–6, 2011.

[3] W. Yuan, K. Dana, M. Varga, A. Ashok, M. Gruteser, and N. Mandayam, "Computer vision methods for visual mimo optical systems," *Proceedings of the IEEE International Workshop on Projector-Camera Systems (held with CVPR)*, pp. 37–43, 2011.

[4] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727 – 752, 2010.

[5] P. Wayner, *Disappearing Cryptography: Information Hiding: Steganography & Watermarking*. Morgan Kaufmann Publishers Inc., 2009.

[6] A. Sangeetha, B. Gomathy, and K. Anusudha, "A watermarking approach to combat geometric attacks," in *Digital Image Processing, 2009 International Conference on*, 2009, pp. 381 –385.

[7] J. L. Dugelay, S. Roche, C. Rey, and G. Doerr, "Still-image watermarking robust to local geometric distortions," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2831–2842, 2006.

[8] X. Wang, L. Hou, and J. Wu, "A feature-based robust digital image watermarking against geometric attacks," *Image Vision Comput.*, vol. 26, no. 7, 2008.

[9] F. Zou, H. Ling, X. Li, Z. Xu, and P. Li, "Robust image copy detection using local invariant feature," in *Multimedia Information Networking and Security, 2009. MINES '09. International Conference on*, vol. 1, 2009, pp. 57–61.

[10] L. Yang and Z. Guo, "A robust video watermarking scheme resilient to spatial desynchronization and photometric distortion," in *Signal Processing, 2006 8th International Conference on*, vol. 4, 16-20 2006.

[11] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *ACM SIGGRAPH*, pp. 369–378, 1997.

[12] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 472–479, 2000.

[13] S. J. Kim, H. T. Lin, Z. Lu, S. Susstrunk, S. Lin, and M. S. BrownHai, "A new in-camera imaging model for color computer vision and its application," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[14] H. T. Lin, S. J. Kim, S. Susstrunk, and M. S. Brown, "Revisiting radiometric calibration for color computer vision," *ICCV*, 2011.

[15] Y. Xiong, K. Saenko, T. Darrell, and T. Zickler, "From pixels to physics: Probabilistic color de-rendering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. pp. 358–365, 2012.

[16] K. Kamijo, N. Kamijo, and G. Zhang, "Invisible barcode with optimized error correction," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, oct. 2008, pp. 2036 –2039.

[17] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar, "Bokode: imperceptible visual tags for camera based interaction from a distance," in *SIGGRAPH*. ACM, 2009.

[18] S. D. Perli, N. Ahmed, and D. Katabi, "Pixnet:designing interference-free wireless links using lcd-camera pairs," *ACM Int. Conf. on Mobile Computing and Networking*, 2010.

[19] J. Vucic, C. Kottke, S. Nerreter, K. Langer, and J. Walewski, "513 mbit/s visible light communications link based on dmt-modulation of a white led," *Journal of Lightwave Technology*, pp. 3512–3518, 2010.

[20] G. Woo, A. Lippman, and R. Raskar, "Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter," *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, vol. 0, pp. 59–64, 2012.

[21] W. Yuan, K. Dana, A. Ashok, M. Varga, M. Gruteser, and N. Mandayam, "Photographic steganography for visual mimo: A computer vision approach," *IEEE Workshop on the Applications of Computer Vision (WACV)*, pp. 345–352, 2012.