Reading Between the Pixels: Photographic Steganography for Camera Display Messaging

Eric Wengrowski*, Kristin J. Dana†, Marco Gruteser and Narayan Mandayam

Department of Electrical and Computer Engineering and WINLAB, Rutgers University

Abstract

We exploit human color metamers to send light-modulated messages decipherable by cameras, but camouflaged to human vision. These time-varying messages are concealed in ordinary images and videos. Unlike previous methods which rely on visually obtrusive intensity modulation, embedding with color reduces visible artifacts. The mismatch in human and camera spectral sensitivity creates a unique opportunity for hidden messaging. Each color pixel in an electronic display image is modified by shifting the base color along a particular color gradient. The challenge is to find the set of color gradients that maximizes camera response and minimizes human response. Our approach does not require a priori measurement of these sensitivity curves. We learn an ellipsoidal partitioning of the 6-dimensional space of base colors and color gradients. This partitioning creates metamer sets defined by the base color of each display pixel and the corresponding color gradient for message encoding. We sample from the learned metamer sets to find optimal color steps for arbitrary base colors. Ordinary displays and cameras are used, so there is no need for high speed cameras or displays. Our primary contribution is a method to map pixels in an arbitrary image to metamer pairs for steganographic camera-display messaging.

1. Introduction

Electronic displays, such as LCD monitors, are typically used only for human visual observation. Research in the relatively new field of camera-display communication has introduced a dual channel: a machine-readable communications channel operating in parallel with the human-observable display. Time-varying messages can be embedded in the on-screen images, but this task has significant challenges. The modulated signal is an illumination field propagating in free-space, so prior methods of watermarking for digital images are not directly applicable. The illumination field emitted by the display and captured by the camera depends on the parameters of the radiometric transfer function and sensitivity curves of both the display and camera. This camera-display transfer function makes message recovery challenging, but it also presents an opportunity for message embedding that is tuned to typical transfer functions.

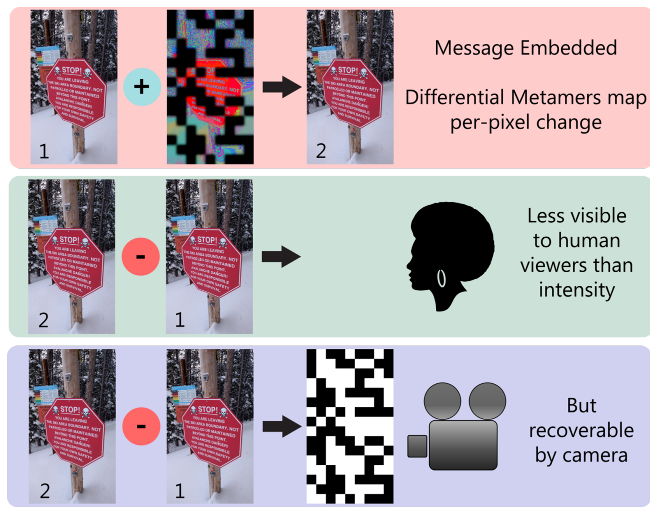


Figure 1: Differential Metamers are color pairs that are optimized to be identical under human vision, but distinguishable by a camera. By modulating small per-pixel changes within an image sequence, differential metamers can be used to embed hidden messages. The embedded message is blended to reduce the spatial visibility without disturbing camera recovery.

A common method for camera-display messaging relies on intensity modulation either for directly embedding bits or for embedding transformation coefficients [1, 2]. Human

*eric.wengrowski@gmail.com
†kdana@ece.rutgers.edu

vision is generally very sensitive to intensity step edges, even when the step size is small. For simple messaging, the display image can be modified by adding a message image where "1" bit values are encoded in a block by a small intensity step and "0" bit values are encoded by zero intensity step. The message frame is added in alternative temporal frames so that sequential frame subtraction can be used to decode the message. This method assumes that the display image is constant over time intervals. Accurate message recovery is challenging because small intensity steps are needed to hide the message, but large intensity steps are needed for a low-noise signal that can be accurately decoded by the camera.

Another approach to making the message imperceptible is to use high speed light modulation so that the flicker fusion effect of human vision can temporally blur the intensity variation [3]. High speed displays are commercially available, but the higher cost is prohibitive for ubiquity in electronic signage and mobile display applications.

Our approach uses *color modulation* that exploits the differences in human color sensitivity versus camera color sensitivity. This allows us to accurately send and receive camouflaged messages without specialized hardware. In a displayed image i, let the pixel coordinate be denoted by $w \in \mathbf{R}^2$. Each image pixel $i(w)$ has 3 color components, $i(w) \in \mathbf{R}^3$. A color message image m is added to i such that our steganographically embedded image $e = i + m$, and each pixel of the embedded message is given by $e(w) = i(w) + m(w)$. For "1" bits, the message m is a color shift added to i. The goal is to find the best color shift $\delta \in \mathbf{R}^3$. Let $\hat{\delta}$ denoted the unit direction in color space, and $\|\delta\|_2$ is the magnitude of the step-size. We seek a step δ to create a *differential metamer* $(i(w), i(w) + \delta)$ such that $i(w) + \delta$ is perceived to be the same color as $i(w)$ by a human observer but is camera-captured as a distinguishable color.

Large sets of differential metamers can be generated given a small training set. Our approach uses a 6-dimensional quadratic binary classifier, solved in a convex optimization problem. Using training data with positive and negative examples, the algorithm determines a set of separating ellipsoids in 6-dimensional space. The interior of these ellipsoids contain 6-dimensional points g where the first three components corresponds to a particular base color and the last three components provide the corresponding $\hat{\delta}$ used for messaging. The interior of these 6-dimensional ellipsoids define approximate metamer sets that sufficiently provide message hiding and recovery.

Differential Metamers

Traditionally, *metamers* are colors that have different spectral power distributions, but appear identical to observers when integrated over the 3 cones sensitivities in the human eye (see Figure 2). We introduce the term *differential metamers* to define pairs of color values programmed for sequential display that result in minimal visible change for the human observer but are distinguishable colors when captured by a camera. This process is illustrated in Figure 1. Many differential metamers exist even among 8-bit color values, but finding the color values that yield both low human sensitivity and high camera sensitivity is difficult because 256^6 (over 2×10^{14}) colors would need to be tested for both camera-display sensitivity and human-display sensitivity. Specific camera sensitivity curves combined with human vision parameters are not be enough to model the differential metamer space. Display parameters indicating the spectrum of light emission for each programmed color vector and the dependence on radiometric observation parameters are also be needed to determine an analytical model. Given the variations involved, we choose a data driven approach instead. We show that this approach is straightforward and effective. We generate samples in 6D space indicating base colors and color gradients for messaging. By observation of the resulting messaging visibility (human and camera), these sample points are labeled as "good" or "bad" for messaging. By sampling 2480 points, we train a set of ellipsoidal binary classifiers that predict successful differential metamers where the base color values i fill the displayable color space. We perform the metamer set estimation in both RGB and CIE Lab color spaces.

2. Background and Related Work

Metamers and Separating Ellipsoids Our approach to finding separating ellipsoids in color space is motivated by two main factors. First, the problem of fitting a separating ellipsoid to labeled data is a convex optimization problem [6] and therefore is not affected by local minima. Second, human vision research has showed the utility of ellipsoidal surface fitting for representing color difference thresholds. As early as the 1940's, human vision studies identified and quantified ellipsoidal representations for the problem of understanding human sensitivity to small color differences [7, 4] as illustrated in Figure 2. This ellipsoidal representation has been confirmed in numerous studies in early vision literature [8, 9, 10]. Parametric surfaces were used to find discriminating contours. The fitting typically used detection thresholds [11, 12] in order to get just-noticeable-difference (JND) contours [13]. Our framework greatly simplifies this process because no threshold values are measured. Metamer sets [14] are convex hulls, which ellipsoids are well-suited to fit. By extension, we have adopted discriminating ellipsoids to characterize the space of differential metamers. In prior work that used color to embed information [15] color gradients are used to watermark spatially varying microstructures into images. The objective in this work is to embed watermarks that were difficult to see from a distance, but visible up close. This is different from our

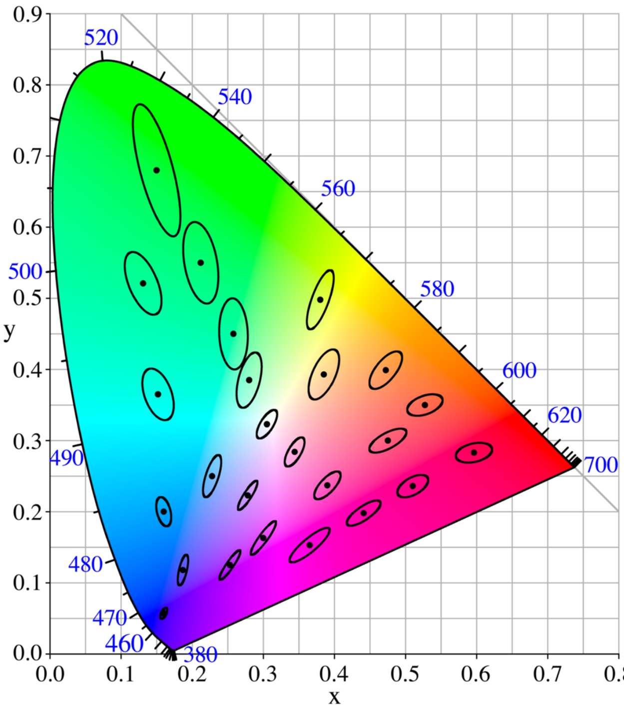


Figure 2: MacAdam ellipses for the CIE xy 1931 colorspace [4, 5]. The area within these scaled-up ellipses represent metamers, or colors which cannot be distinguished.

goal of finding pairs of colors where no distinction can be made when viewed sequentially by humans, but the difference can be robustly detected by a camera.

Camera-Display Communication Electronic displays such as televisions, computer monitors, and projectors are traditionally used to display images, videos, and text - all human readable scenes. These devices can also display camera-readable images such as QR-codes [1, 16, 17, 18, 19, 20, 21, 22, 2, 23, 24, 25, 26, 27]. Within the past 5 years, extensive work has been done to expand the capabilities of camera display messaging by increasing throughput.

PixNet introduced OFDM transmission algorithms to address the unique characteristics of the camera-display link, including perspective distortion, blur, and sensitivity to ambient light [22]. While PixNet offer impressive data throughput, it can only display machine-readable code and supports no hybrid approach. Strata introduced distance-scalable coding schemes [16], preferable in a mobile application, but also cannot display both human-readable and camera readable images at the same time. Both of the aforementioned techniques encode bit values with intensity. COBRA introduced a 2D color code [17], but also could only display machine readable code.

Both Visual MIMO [1, 19, 20, 24, 25, 26, 27] and HiLight [21] use intensity modulation in human-readable images to embed a second machine-readable channel. However, it is well known that human vision is extremely sensitive to temporal and spatial changes in intensity. It has been shown that intensity changes, even with small magnitude are likely to cause flicker and discomfort to a human observer. The amount of human visual obtrusion had not been measured for either method.

Kaleido [28] and VRCodes [29] uses metamers to embed data in alternating pixel values. These values, however, are not “true” metamers in the sense that two static colors have different physical properties such as wavelength, but appear identical to human viewers. Instead, Kaleido and VRCodes leverage flicker fusion to create temporally blended colors hidden from human observers with high speed changes. This approach is constrained by the need for specialized high-speed displays and cameras. VRCodes also leverages the rolling shutter camera typically found on mobile phones to sample at frequencies above 60Hz. Unfortunately, this limits VRCode throughput to only 1 bit per frame.

Kaleido [28] attempts to solve a different problem: embedding noise with flicker fusion metamers to disrupt piracy via camera recording of videos, while preserving the human-visible channel. While similar in intuition to the work presented in this paper, the goals are fundamentally different. We embed camera-sensitive information in this invisible channel, while Kaleido only embeds camera-sensitive noise. And as stated before, Kaleido requires specialized high-speed displays, while our method requires no specialized hardware.

LED arrays have used modulated light to communicate [25, 30, 23]. Recently, LED-based communication techniques have used color-shift keying for communication [31]. Methods exist to make this color-shift keying imperceptible to human observers [32], but these applications do not require the imperceptible reproduction of high resolution images.

In this work, we take a data driven approach to generating differential metamers that have a small human sensitivity gradient, but large camera sensitivity gradient. We show that differential metamers are effective for steganographically embedding messages into high-quality images on electronic displays.

3. Photographic Steganography System Design

Embedding Steganographic Messages The message structure we employ is a 2D barcode grid, 16 blocks wide and 9 blocks tall, containing 144 bits in total. The barcode spans the entire display area. To reduce the visible artifacts from sharp spatial gradients, the block pattern is blended. The dimensions of the 2D barcode were chosen empirically. With smaller blocks, more bits can be transmitted in a single image. But as spatial redundancy is reduced, bit recovery errors will increase. Messages larger then 144 bits can be constructed by stringing together sequential 144-bit messages. For each block, a color shift keys a "1" bit. No change to the base color keys a "0" bit.

We represent a differential metamer as the 6-dimensional vector g separated into two components $g = [g_b \ g_m]^T$ where g_b is the base color in Lab space with $g_b \in \mathbf{R}^3$ and g_m is the optimal color shift $\delta \in \mathbf{R}^3$ in the same color space.

The core problem is finding the optimal δ for an arbitrary pixel base color. We denote G as a set of differential metamers. For each pixel coordinate w, we compute the minimum distance between $i(w)$ and g_b for every member of G. We refer to the g with the nearest g_b as g^*, and g^*_m provides the corresponding color shift for $i(w)$. So if $i(w)$ belongs to a block keyed with a "1" bit, then $e(w) = i(w) + \delta$.

When the images i and e are rendered, they are transformed by the display's spectral emittance function $D()$ which is unknown. When the images are displayed in a video sequence, odd frames display the original image $D(i)$, and even frames display the steganographically embedded image $D(e)$.

Recovering Steganographic Messages The two image frames are sequentially imaged by the camera. The displayed images are affected by light travel in free space and are transformed by the camera's spectral sensitivity function. Denote these two unknown transformation functions $F()$ and $C()$ respectively. The camera-captured images $C(F(D(i)))$ and $C(F(D(e)))$ are subtracted from each other. For each bit-block, an average difference greater than some threshold corresponds to a "1", and below that threshold corresponds to a "0". The threshold is calculated by reserving 4 of the 144 bits for calibration. The recovered message was then compared to the known message to calculate BER (bit error rate). BER is the percentage of misclassified bits in each 144 bit message.

$$\text{BER} = \frac{\text{count(incorrectly classified bits)}}{\text{count(all bits)}},$$

4. Learning New Differential Metamers

As stated in Section 1, differential metamers exist even among 8-bit color values. But testing 256^6 colors is expensive and undesirable. Our approach for generating an expanded gamut of differential metamers relies on a training set of base colors $i(w)$ and color shift gradients δ. Positive examples in this training set meet the criteria for embedding: no visible flicker and accurate camera recovery. Negative examples do not meet the criteria for embedding: color pairs that are either visible when viewed sequentially or not recoverable by the camera.

The data resides in 6-dimensional space $\mathbf{R}^6$. We choose the number of separating ellipsoids k empirically and cluster the positive examples into k clusters in $\mathbf{R}^6$. For each cluster, we use convex optimization to find the optimal ellipsoid that separates positive from negative data. Sampling within the union of all separating ellipsoids reveals a dense set of new differential metamers.

For each cluster k_i, the optimal separating ellipsoid is found. Each ellipsoid separates the positive training examples in cluster k_i from all negative training examples.

Collecting and Labeling Training Data

The set of 124 base colors are generated by uniformly sampling CIE Lab space. For each base color, 20 baricentrically sampled unit vectors are generated. In total, we now have 2480 training examples.

A video sequence is generated. Odd frames consist of only a monochromatic image of the base color. Even frames comprise the base color plus a 2D barcode grid corresponding to a message. The magnitude of the color step size is defined as the L-2 Norm:

$$\|\delta\|_2 = \sqrt{\delta_L^2 + \delta_a^2 + \delta_b^2}$$

Here, $\|\delta\|_2 = 5$ (in the 8-bit [0 255] scale). For these tests, the same checkerboard message is used every time, since it maximizes spatial variation and is likely to be noticed by humans. A camera views the 2480 image sequences only once and attempts to recover the embedded messages. The camera is fixed 0.5 meters from the display with a viewing angle normal to the image plane.

For each of the 2480 training examples, human participants were shown video sequences each containing a single color and with an embedded checkerboard pattern alternating at 8Hz for 10 seconds. 8Hz was chosen because humans are particularly sensitive to intensity changes at this frequency [33], and because it represents a reasonable target for smartphone video capture rates. The participants were asked to indicate if they could see the checkerboard pattern or not. Three participants were used for human vision evaluation. They were students between ages 19 and 24. One participant wore glasses, and none had any color-blindness. The variance in their flicker labeling was negligible.

Single color images are used to isolate the exact behavior of each color pair, and negating the cloaking effects of image content (e.g. texture) and preventing participants from confusing the effects of other, nearby pixels. Relative contrast may have an effect on visibility in real images, but this can be overcome by embedding differential metamers only in a select subset of pixels, or by first clustering nearby pixels by differential metamer gradients and not embedding on the cluster borders. While an evaluation of spatial obtrusiveness caused by relative contrast is interesting, it is outside the scope of this paper and left for future work.

Positive training examples are defined as ones whose color embedding were completely invisible to humans, but

Figure 3: A separating ellipsoid and single cluster of positively labeled training data. Visualizing k 6-dimensional ellipsoids is difficult, so the data has been projected down to 2D a b space (from Lab color space). We show base colors and color shifts at the same time. The solid circle represents the base colors i, and its respective line segments represents the color shift gradients δ. The color of each circle and line segment is the actual base color. Notice how there is a general axis of color shift direction for the data in this cluster. Since these are positive training examples, this indicates that human viewers are relatively insensitive to these color shifts. This also indicates that our camera is sensitive to these color shifts.

recoverable by camera with BER (bit error rate) $= 0\%$. All other examples were labeled negative training data. After labeling, 922 positive and 1558 negative examples were used for training. Empirically, we chose the number of clusters $k = 50$.

Learning k Optimal Separating Ellipsoids

We have two sets of points in $\mathbf{R}^6$, $\{x_1, ...x_N\}$ and $\{y_1, ...y_M\}$. The points x_i represent the base colors and modulation steps that satisfy the requirements for embedding: BER $= 0\%$, and no visible flicker. While the points y_i do not satisfy both of these conditions. We wish to find a function $f : \mathbf{R}^n \to \mathbf{R}$ that is positive on the first set, and negative on the second, *i.e.*,

$$f(x_i) > 0, \; i = 1, ..., N, \qquad f(y_i) < 0, \; i = 1, ..., M. \tag{1}$$

When these inequalities hold, we say that f separates the two sets of points.

Quadratic Discrimination Since our data points cannot be separated by a N-dimensional hyperplane, we seek classification via nonlinear discrimination. As long as the parameters that define f are linear (or affine), the above inequality can still be solved with convex optimization.

In this case, we choose f to be quadratic and in homogeneous form:

$$f(z) = z^T P z + q^T z + r, \tag{2}$$

where $P \in \mathbf{S}^n$ (P is a symmetric $n \times n$ matrix), $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$, with dimensionality n = 6. Those parameters $P, \; q, \; r$ are bound by the following constraints:

$$\begin{aligned} x_i^T P x_i + q^T x_i + r &> 0, \qquad && i = 1, ..., N \\ y_i^T P y_i + q^T y_i + r &< 0, \qquad && i = 1, ..., M \end{aligned} \tag{3}$$

Next, we replace 0 with ϵ, creating a separating band that is 2ϵ wide:

$$\begin{aligned} x_i^T P x_i + q^T x_i + r &\geq \epsilon, \qquad && i = 1, ..., N \\ y_i^T P y_i + q^T y_i + r &\leq -\epsilon, \qquad && i = 1, ..., M \end{aligned} \tag{4}$$

Dividing out by ϵ and subsuming the scalar $\frac{1}{\epsilon}$ into P, q, r, you arrive at Eq. 5. Following [6], we solve for the parameters $P, \; q, \; r$ by solving the non-strict feasibility problem:

$$\begin{aligned} x_i^T P x_i + q^T x_i + r &\geq 1, \qquad && i = 1, ..., N \\ y_i^T P y_i + q^T y_i + r &\leq -1, \qquad && i = 1, ..., M \end{aligned} \tag{5}$$

The resulting separating surface $\{z \mid z^T P z + q^T z + r = 0\}$ is quadratic.

Separating Ellipsoids We can change the shape of our quadratic separating surface by imposing additional constraints on the parameters P, q, and r. We form an ellipsoid that contains all points $x_i, ..., x_N$ but none of the points $y_i, ..., y_M$ by requiring that $P \prec 0$, that is P is negative definite. We can use homogeneity in $P, \; q, \; r$ to express the constraint $P \prec 0$ as $P \preceq -I$. We can then cast our quadratic discrimination problem as the following semi-define programming (SDP) feasibility problem:

$$\begin{array}{lll} \text{find} & P, \; q, \; r & \\ \text{subject to} & x_i^T P x_i + q^T x_i + r \geq 1, & i = 1, ..., N \\ & y_i^T P y_i + q^T y_i + r \leq -1, & i = 1, ..., M \\ & P \preceq -I & \end{array} \tag{6}$$

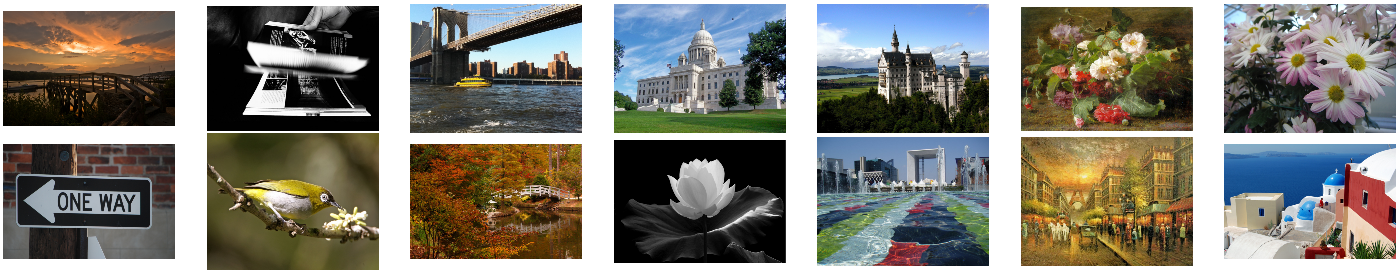


Figure 4: Set of 14 images used to evaluate BER across several embedding algorithms and message step-sizes.

While technically correct, this optimization problem will fail if any of the training points fall outside their classification boundaries. Following the development in [6] for support vector classifiers, we relax our constraints by introducing non-negative variables $u_1, ..., u_N$ and $v_1, ..., v_M$. With the relaxation variables u_i and v_i introduced, our inequalities become:

$$\begin{aligned} x_i^T P x_i + q^T x_i + r &\geq 1 - u_i, & i &= 1, ..., N \\ y_i^T P y_i + q^T y_i + r &\leq v_i - 1, & i &= 1, ..., M \end{aligned} \tag{7}$$

The relaxation variables u_i and v_i represent the distances of each point outside it's proper boundary. In the original problem, $u = v = 0$. We can think of u_i as a measure of how much each constraint $x_i^T P x_i + q^T x_i + r \geq 1$ is being violated and that's what we want to minimize. A good heuristic is minimizing the sum of variables u_i and v_i. The separating ellipsoid defined by $P,\ q,\ r$ is found with the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \mathbf{1}^T u + \mathbf{1}^T v \\ \text{subject to} \quad & x_i^T P x_i + q^T x_i + r \geq 1 - u_i, \qquad i = 1, ..., N \\ & y_i^T P y_i + q^T y_i + r \leq v_i - 1, \qquad i = 1, ..., M \\ & P \preceq -I \\ & u \preceq 0, \qquad v \preceq 0 \end{aligned} \tag{8}$$

To solve this problem we used CVX, a package for specifying and solving convex programs [34, 35]. After each ellipsoid is solved, we test that the ellipsoid is populated before accepting it.

Sampling Within Union of Ellipsoids Once k optimal separating ellipsoids are trained, the points inside the ellipsoids reflect desirable values for message embedding. So to expand our gamut of differential metamers, we densely sample inside the ellipsoid region for new points. G' is the expanded set of newly generated differential metamers g'.

5. Experiments

We wish to evaluate the expanded set of differential metamers learned using the techniques described in Section 4. For each of our embedding algorithms, a known message was embedded into a pair of 2 images. A camera then sequentially captured the original image, then the image with the embedded message pattern. Again, the camera was a Basler acA2040-90uc-CVM4000, and the display was an Acer S240HL IPS LCD monitor. The camera was stationed approximately 0.5 meters from the electronic display. The camera had a fixed shutter speed, ISO sensitivity, aperture, and white balance. Each algorithm was evaluated based on the accuracy of recovering each bit of the message. A wide range of message step-sizes were tested. Message step-size refers to the $\|\delta\|_2$, or δ magnitude in 8-bit pixel values. A diverse set of 14 host images was used, shown in Table 4.

For the intensity-based approach, a uniform grayscale δ is applied to every pixel representing a "1" bit. The random approach applies a δ in a random direction to each pixel. The RGB differential metamers approach assigns a specialized δ value to each pixel in the base image. The differential metamer ellipsoids are trained in 6-dimensional RGB space. Similarly, the *Lab* differential metamers approach assigns δ values from ellipsoids trained in 6-dimensional *Lab* space.

Evaluation of Clustering Methods

A series of clustering algorithms were evaluated: kmeans, kmediods, Gaussian Mixture Models, Hierarchical clustering, and Spectral clustering. Ellipsoids were trained and learned using each of these clustering methods. The ellipsoids yielded differential metamers used for steganographic embedding and recovery. This evaluation is performed twice for each clustering algorithm under two different illumination conditions. Once where the camera has fixed high-exposure settings, and once again with fixed low-exposure settings.

The respective mean errors were 27.285%, 25.97%, 25.025%, 26.17%, and 29.61%. The respective run times were 0.0435s, 0.5813s, 0.0978s, 0.1299s, and 0.1387s.

	Embedding Algorithms			
$\|\|\delta\|\|_2$	Intensity	Random	RGB Differential Metamers	Lab Differential Metamers
1	50.69%	50.99%	50.45%	49.85%
2	47.92%	48.81%	42.06%	42.06%
3	43.85%	46.97%	36.11%	37.25%
4	37.00%	44.59%	29.02%	27.83%
5	34.52%	42.41%	22.42%	21.73%
6	23.41%	41.22%	19.84%	17.61%
7	18.70%	38.10%	15.53%	15.08%
8	13.49%	35.57%	13.84%	12.80%
9	09.97%	34.72%	12.50%	12.00%
10	09.13%	32.89%	11.01%	10.91%

Table 1: BER for various embedding schemes (*lower is better*). The red-shaded cells indicate δ magnitudes where an blended message pattern is easily visible. The green-shaded cells indicate optimal values where the blended message pattern is camouflaged from human vision, but in a good position to be camera-recovered. Differential metamers generated with trained ellipsoids in CIE Lab are especially effective because both the BER is reduced and the threshold for acceptable step-size is increased. Notice that for a mid-range step-size of 5 or 6, the Lab differential metamers significantly outperform intensity modulation.

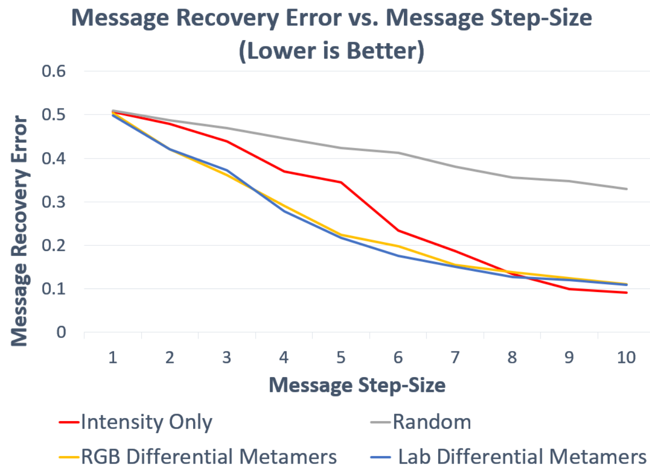


Figure 5: This graph compares message recovery across several embedding algorithms. Regardless of embedding algorithm, as message step size ($||\delta||_2$) increases, message recovery error decreases. However, large step size also means a more visually obtrusive embedding. For an embedded message to be invisible, smaller step size are greatly preferred. For small to mid-range $||\delta||_2$, color embedding using differential metamers is significantly better.

Gaussian Mixture Models (GMMs) yielded the lowest BER on average. Although the margin of superiority was small, Gaussian mixture models were chosen as the best balance of error and run-time. Regardless of method used or illumination condition, the standard deviation hovered around 10% for all methods. This suggests that the recovery error results are largely dependent on the base image used. This result has been verified empirically as well; certain images produce better embedding results. The run time calculations took place on an Intel 6700K processor with 32 GB of memory running Matlab 2015b.

6. Results

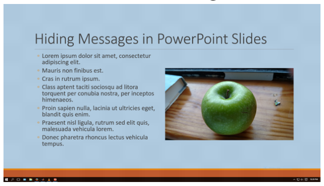
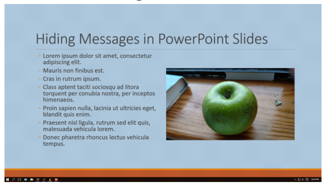
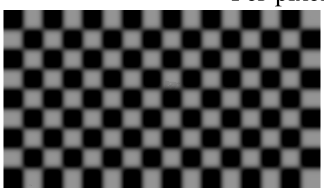
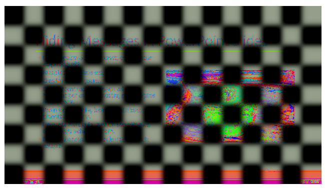
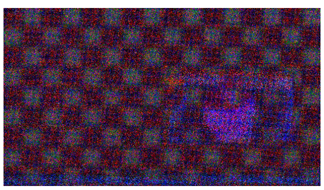
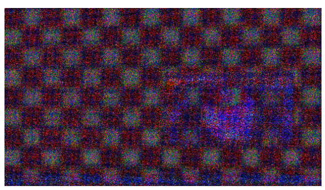
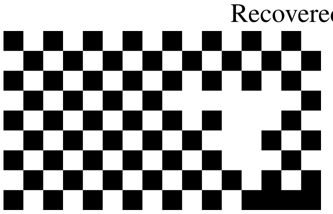
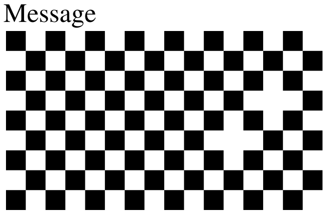
Table 1 shows the average message recovery for each embedding algorithm across a variety of $||\delta||_2$ values (step sizes). The red-shaded cells represent values for which the $||\delta||_2$ is so large, the message pattern can be obviously detected by humans. Figure 5 illustrates these results graphically.

For small $||\delta||_2$, the RGB and Lab differential metamer approaches greatly outperform the alternatives. Small step sizes are typically preferable because they are more difficult for humans to see. With the differential metamer approach, larger step size can be used, facilitating more accurate camera recovery. The differential metamers trained in Lab space are most effective at reducing human detection with most robust message recovery. Table 2 illustrates these results.

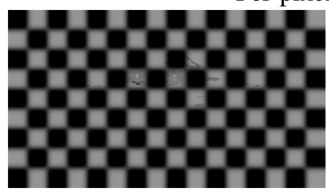
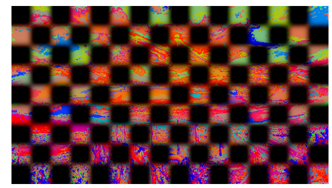
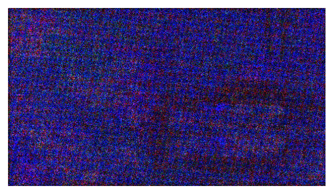
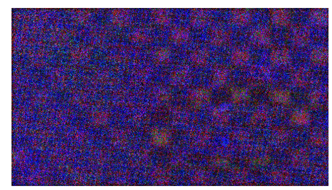
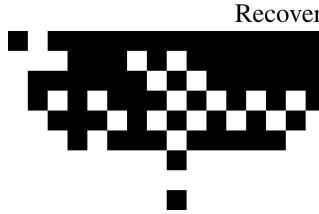
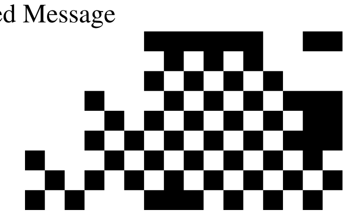
Although the mean error is high compared to perfect recovery, it can be functionally reduced using error-correcting codes. The proposed color messaging framework is applicable to more sophisticated photo-steganographic messaging systems. For the purposes of this paper, only the reduction in error due to color messaging is evaluated.

Transferring Learned Ellipsoids to New Hardware

The results presented thus far showcase the effectiveness of photographic steganography using differential metamers trained on a single camera-display pair. But we want to know how well our learned ellipsoids will transfer to a new camera-display pair. If new differential metamers must be learned for every camera-display combination, the applicability of our algorithm is limited. Table 3 features experimental results when the camera-display pair used for photographic steganography is totally different from the camera-display pair used for training. Although the illumination conditions and imaging pipeline remain unchanged, the most significant aspects of the system have been changed. When using different hardware, the BER increases by only 3.48%. Using the same hardware, transfered differential metamers significantly outperform intensity-based embed-

Photographic Steganography: Intensity vs CIE Lab Differential Metamers	
Intensity	Differential Metamers
Image with Embedded Message	
	
Per-pixel difference	
	
Camera-recovered difference	
	
Recovered Message	
	
BER *(lower is better)*	
5.56%	1.39%

(a) Low Texture Image

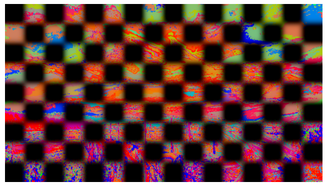
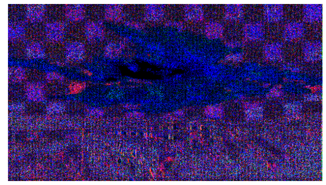
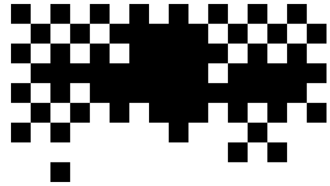
Photographic Steganography: Intensity vs CIE Lab Differential Metamers	
Intensity	Differential Metamers
Image with Embedded Message	
	
Per-pixel difference	
	
Camera-recovered difference	
	
Recovered Message	
	
BER *(lower is better)*	
38.19%	19.44%

(b) Highly Textured Image

Table 2: Message embedding with intensity vs differential metamers example. The image in the first row contains a steganographic message pattern. Below that, the per-pixel difference shows the ground truth of exactly the changes that were made to the original image. The camera-recovered difference shows the difference measured after the image has been displayed electronically, and captured by a camera. Notice that the differences between ground truth and camera-captured are large. Embedding messages with Lab differential metamers is effective for many types of images, including slide or sign type images, as is shown in (a). The example in (b) showcases a more challenging natural image case, where intensity embedding fails in dark and highly textured areas of the image. Lab differential metamers are significantly more effective for robust message embedding and recovery. In both (a) and (b), $\|\delta\|_2 = 5$ for all algorithms.

ding. The differential metamers learned under certain hardware conditions can be transfered for a small accuracy cost.

Transferring Learned Ellipsoids to a New Camera-Display Pair

Transfered Differential Metamers			
Image with Embedded Message	Per-pixel difference	Camera-recovered difference	Recovered Message
			
Transfered Differential Metamers BER $= 22.92\%$ *(lower is better)*			

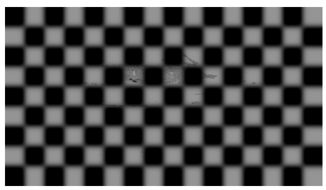
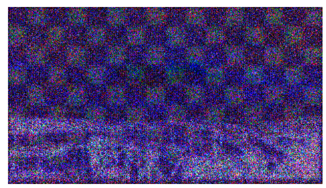
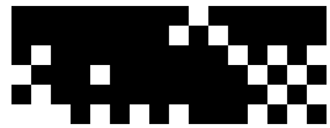
Intensity Embedding			
Image with Embedded Message	Per-pixel difference	Camera-recovered difference	Recovered Message
			
Intensity BER $= 36.11\%$ *(lower is better)*			

Table 3: Photographic Steganography using differential metamers learned with a different camera-display pair. An Acer Predator MNT XB271HUC IPS display and Basler acA1300-30uc camera were used in experiment. However, the ellipsoids yielding differential metamers were trained using the aforementioned Basler acA2040-90uc-CVM4000 camera and Acer S240HL display. With $\|\delta\|_2 = 5$, the recovered message has a BER of 22.92%, only 3.48% worse than the hardware used for training as shown in Table 2. This example demonstrates that the ellipsoids learned can be robustly transfered between different hardware and still significantly outperform intensity-based embedding.

7. Discussion and Conclusion

In this paper, we present a color modulation method used to steganographically embed messages into ordinary images and videos. We develop a data-driven approach to learn a pixel mapping function that produces an optimal differential metamer pair for any pixel value. These differential metamers are pairs of color values that minimize human visual response, but maximize camera response. The key innovation is a novel color-selection framework that leverages the mismatch between human spectral and camera sensitivity curves. We refer to this task of camouflaged camera-display messaging as photographic steganography.

We demonstrate the effectiveness of our differential metamer generation algorithm with message embedding. The goal is to maximize throughput, minimize recovery error, and camouflage the visible artifacts to humans. The desirability of our approach stems from the creation of a communication side-channel without using specialized hardware. Embedded information could be used to grant access that is *conditioned on close physical presence* (for security or convenience). Unlike NFC (near-field communications) which is commonly used for precise location verification but has problems with network saturation for nodes in close proximity, beacons using photographic steganography would ensure that users are facing a particular direction. For example, users would not be able to access a networked projector unless they used photographic steganography to recover a dynamic access code embedded in the projectors displayed images to prove that they are in the appropriate location. Scenarios include those where users perform scavenger-hunt games in museums or use outdoor electronic billboards for tickets/coupons/schedules. It is also easy to envision a scenario where users install a smartphone application and have access to extra content on live-broadcast videos.

Acknowledgements

This work was supported primarily by the National Science Foundation under NSF grant CNS-1065463. The Titan X used for this research was donated by the NVIDIA Corporation. We thank Peter Meer for valuable discussions on metamers in human vision.

References

[1] Eric Wengrowski, Wenjia Yuan, Kristin J Dana, Ashwin Ashok, Marco Gruteser, and Narayan Mandayam. Optimal radiometric calibration for camera-display communication. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016.

[2] Kensei Jo, Mohit Gupta, and Shree K Nayar. Disco: Display-camera communication using rolling shutter sensors. *ACM Transactions on Graphics (TOG)*, 35(5):150, 2016.

[3] Viet Nguyen, Yaqin Tang, Ashwin Ashok, Marco Gruteser, Kristin Dana, Wenjun Hu, Eric Wengrowski, and Narayan Mandayam. High–rate flicker–free screen–camera communication with spatially adaptive embedding. In *IEEE INFOCOM*, volume 2, 2016.

[4] David L MacAdam. Specification of small chromaticity differences. *JOSA*, 33(1):18–26, 1943.

[5] Wikimedia Commons. Ciexy1931 macadam.png. `https://en.wikipedia.org/wiki/File:CIExy1931_MacAdam.png`, 2015.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] WRT Brown and DL MacAdam. Visual sensitivities to combined chromaticity and luminance differences. *JOSA*, 39(10):808–823, 1949.

[8] Günter Wyszecki and GH Fielder. New color-matching ellipses. *JOSA*, 61(9):1135–1152, 1971.

[9] Walter RJ Brown. Color discrimination of twelve observers. *JOSA*, 47(2):137–143, 1957.

[10] Hugh R Davidson. Calculation of color differences from visual sensitivity ellipsoids. *JOSA*, 41(12):1052–1055, 1951.

[11] Allen B. Poirson, Brian A. Wandell, Denise C. Varner, and David H. Brainard. Surface characterizations of color thresholds. *J. Opt. Soc. Am. A*, 7(4):783–789, Apr 1990.

[12] Günter Wyszecki, VS Stiles, and Kenneth L Kelly. Color science: concepts and methods, quantitative data and formulas. *Physics Today*, 21(6):83–84, 2009.

[13] C. Noorlander and J. J. Koenderink. Spatial and temporal discrimination ellipsoids in color space. *J. Opt. Soc. Am.*, 73(11):1533–1543, Nov 1983.

[14] Graham D Finlayson and Peter Morovic. Metamer sets. *JOSA A*, 22(5):810–819, 2005.

[15] Nicolas Rudaz and Roger D Hersch. Protecting identity documents by microstructure color differences. *Journal of Electronic Imaging*, 13(2):315–323, 2004.

[16] Wenjun Hu, Jingshu Mao, Zihui Huang, Yiqing Xue, Junfeng She, Kaigui Bian, and Guobin Shen. Strata: layered coding for scalable visual communication. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 79–90. ACM, 2014.

[17] Tian Hao, Ruogu Zhou, and Guoliang Xing. Cobra: color barcode streaming for smartphone systems. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 85–98. ACM, 2012.

[18] Wenjia Yuan, Richard E Howard, Kristin J Dana, Ramesh Raskar, Ashwin Ashok, Marco Gruteser, and Narayan Mandayam. Phase messaging method for time-of-flight cameras. In *Computational Photography (ICCP), 2014 IEEE International Conference on*, pages 1–8. IEEE, 2014.

[19] Wenjia Yuan, Kristin Dana, Ashwin Ashok, Marco Gruteser, and Narayan Mandayam. Dynamic and invisible messaging for visual mimo. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 345–352. IEEE, 2012.

[20] Wenjia Yuan, Kristin Dana, Michael Varga, Ashwin Ashok, Marco Gruteser, and Narayan Mandayam. Computer vision methods for visual mimo optical system. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 37–43. IEEE, 2011.

[21] Tianxing Li, Chuankai An, Andrew Campbell, and Xia Zhou. Hilight: hiding bits in pixel translucency changes. In *Proceedings of the 1st ACM MobiCom workshop on Visible light communication systems*, pages 45–50. ACM, 2014.

[22] Samuel David Perli, Nabeel Ahmed, and Dina Katabi. Pixnet: interference-free wireless links using lcd-camera pairs. In *Proceedings of the sixteenth an-*

nual international conference on Mobile computing and networking, pages 137–148. ACM, 2010.

[23] Ashwin Ashok, Shubham Jain, Marco Gruteser, Narayan Mandayam, Wenjia Yuan, and Kristin Dana. Capacity of screen–camera communications under perspective distortions. *Pervasive and Mobile Computing*, 16:239–250, 2015.

[24] A. Ashok, M. Gruteser, N. Mandayam, T. Kwon, W. Yuan, M. Varga, and K. Dana. Rate adaptation in visual mimo. *Proceedings of IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pages pp. 583–591, 2011.

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

[26] 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*, pages pp.105–112, 2010.

[27] Ashwin Ashok, Marco Gruteser, Narayan Mandayam, Jayant Silva, Michael Varga, and Kristin Dana. Challenge: mobile optical networks through visual mimo. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, pages 105–112. ACM, 2010.

[28] Lan Zhang, Cheng Bo, Jiahui Hou, Xiang-Yang Li, Yu Wang, Kebin Liu, and Yunhao Liu. Kaleido: You can watch it but cannot record it. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 372–385. ACM, 2015.

[29] Grace Woo, Andrew Lippman, and Ramesh Raskar. Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 59–64. IEEE, 2012.

[30] Ashwin Ashok, Viet Nguyen, Marco Gruteser, Narayan Mandayam, Wenjia Yuan, and Kristin Dana. Do not share!: invisible light beacons for signaling preferences to privacy-respecting cameras. In *Proceedings of the 1st ACM MobiCom workshop on Visible light communication systems*, pages 39–44. ACM, 2014.

[31] Pengfei Luo, Min Zhang, Zabih Ghassemlooy, Hoa Le Minh, Hsin-Mu Tsai, Xuan Tang, Lih Png, and Dahai Han. Experimental demonstration of rgb led-based optical camera communications.

[32] Pengfei Hu, Parth H Pathak, Xiaotao Feng, Hao Fu, and Prasant Mohapatra. Colorbars: Increasing data rate of led-to-camera communication using color shift keying.

[33] DH Kelly. Sine waves and flicker fusion. *Documenta Ophthalmologica*, 18(1):16–35, 1964.

[34] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[35] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.