A

DisCo: Displays that Communicate

Kensei Jo, Columbia University
Mohit Gupta, Columbia University
Shree K. Nayar, Columbia University

We present DisCo, a novel display-camera communication system. DisCo enables displays and cameras to communicate with each other, while also displaying and capturing images for human consumption. Messages are transmitted by temporally modulating the display brightness at high frequencies so that they are imperceptible to humans. Messages are received by a rolling shutter camera which converts the temporally modulated incident light into a spatial flicker pattern. In the captured image, the flicker pattern is superimposed on the pattern shown on the display. The flicker and the display pattern are separated by capturing two images with different exposures. The proposed system performs robustly in challenging real-world situations such as occlusion, variable display size, defocus blur, perspective distortion and camera rotation. Unlike several existing visible light communication methods, DisCo works with off-the-shelf image sensors. It is compatible with a variety of sources (including displays, single LEDs), as well as reflective surfaces illuminated with light sources. We have built hardware prototypes that demonstrate DisCo's performance in several scenarios. Because of its robustness, speed, ease of use and generality, DisCo can be widely deployed in several CHI applications, such as advertising, pairing of displays with cell-phones, tagging objects in stores and museums, and indoor navigation.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - Input devices and strategies

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Image sensor communication, Optical wireless communication, user interfaces

ACM Reference Format:
Kensei Jo, Mohit Gupta and Shree K. Nayar, 2014. DisCo: Displays that Communicate. *Columbia University Technical Report* V, N, Article A (December 16 2014), 20 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

We present DisCo, a novel display-camera communication system that enables displays to send short messages to digital sensors, while simultaneously displaying images for human consumption (Figure 1). Existing display-camera communication methods are largely based on spatial-domain steganography, where the information is encoded as an imperceptible spatial signal (e.g., QR-code). These methods, while simple to implement, are prone to errors due to common causes of image degradations such as occlusions, display being outside the sensor's field-of-view, defocus blur and perspective distortion. Due to these limitations, steganography based techniques have not been widely adopted, specially in uncontrolled settings involving consumer cameras and public displays.

DisCo overcomes these limitations by embedding messages in temporal signals instead of spatial signals. We draw inspiration from the emerging field of visible light communication (VLC), where information is transmitted between a light source (transmitter) and a sensor (receiver) via high frequency temporally modulated light. Most of these techniques require specialized high-speed cameras or photo-diodes as signal receivers [Elgala et al. 2009; Vucic et al. 2010; Sarkera et al. 2009]. Recently, a method was proposed for using low cost rolling shutter sensors as receivers. This method, however, places strong restrictions on the transmitter; only light sources (e.g., LEDs) or

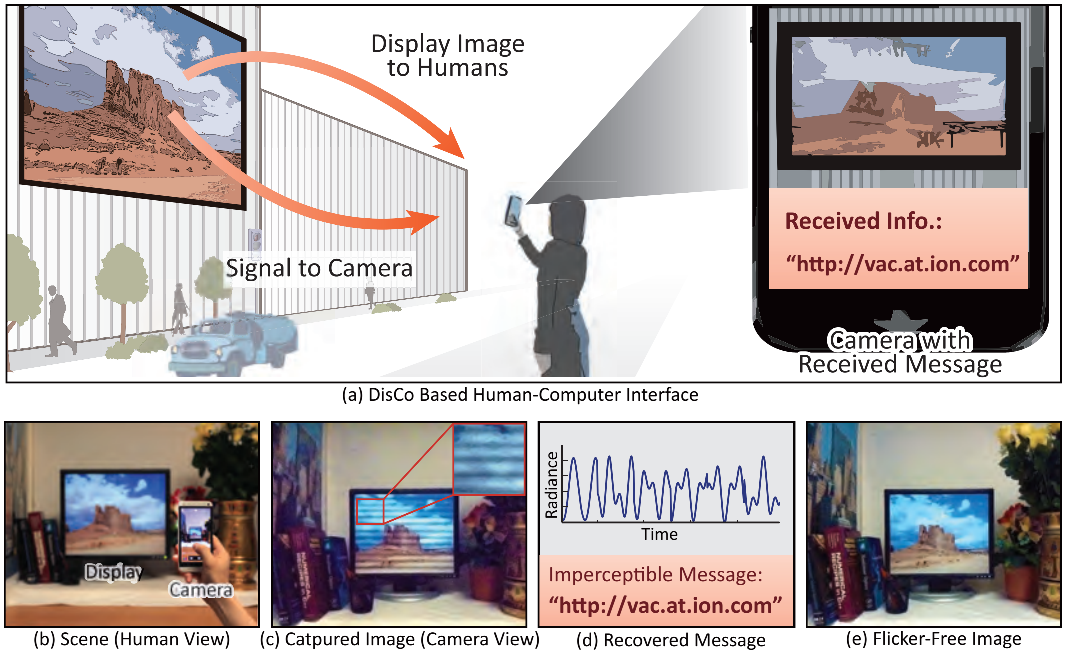


Fig. 1. **Concept of DisCo:** (a) We propose DisCo, a novel human computer interface based on display sensor communication. It uses fast temporal modulation of displays to transmit messages, and rolling shutter sensors to receive them. The messages are imperceptible to humans, allowing displays to serve the dual purposes of displaying images to humans while simultaneously conveying messages to cameras. (b) A scene comprising a display. (c) Image captured by a rolling shutter camera. Due to rolling shutter, temporal modulation of the display is converted into a spatial flicker pattern. The flicker pattern is superimposed on the displayed pattern. By using a sensor that can capture two exposures simultaneously, we can separate the flicker and the display pattern, and thus recover both the message (d), and flicker-free scene image (e) from a single captured image.

surfaces with constant brightness [Danakis et al. 2012] can be used. These systems do not work with displays that need to display arbitrary images. The goal of this paper is on designing systems that can use a broad range of signal transmitters, specially displays showing arbitrary images, as well as objects that are illuminated with temporally modulated light. The objects can have arbitrary textures. This is shown in Figure 2.

DisCo builds upon the method proposed in [Danakis et al. 2012] and uses rolling shutter cameras as signal receivers. In rolling shutter sensors, different rows of pixels are exposed in rapid succession, thereby sampling the incident light at different time instants. This converts the temporally modulated light coming from the display into a spatial *flicker* pattern in the captured image. The flicker encodes the transmitted signal. However, the flicker pattern is superimposed with the (unknown) display pattern. This is illustrated in Figure 1. In order to extract the message, the flicker and the display pattern must be separated. Our *key contribution* is to show that the two components can be separated by capturing images at two different camera exposures. We also show that the flicker component is *invariant* to the display pattern and other common imaging degradations (e.g., defocus blur, occlusion, camera rotation and variable display size). The effect of all these degradations can be absorbed in the display pattern component. Since the display pattern is separated from the flicker component before signal recovery, the imaging degradations do not adversely affect the communication process.

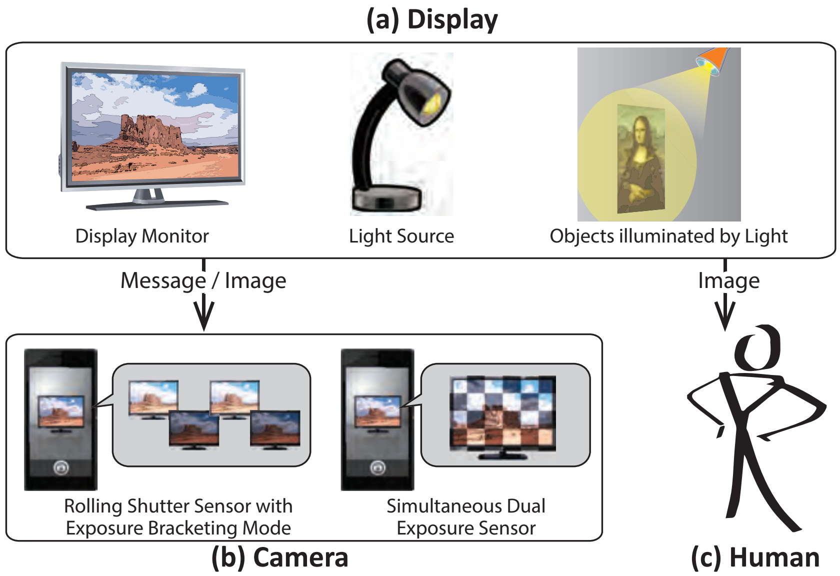


Fig. 2. **Components of DisCo:** DisCo can work with a broad range of devices. (a) For transmitters, DisCo can use display monitors, light sources or objects illuminated by light as the transmitter ("Display"). (b) The display sends information to the receiver, which can be implemented using rolling shutter cameras. (c) The display also shows images for human consumption, while simultaneously communicating with the camera.

Hardware implementation and prototypes: DisCo system consists of two main components - the display (transmitter) and the camera (receiver). The (display) transmitter for DisCo can be implemented as an LCD panel with an LED back-light, or a single LED, or even a non-emitting surface illuminated with a spot-light, as shown in Figure 2. The receiver is a digital camera with a rolling shutter.

We demonstrate two prototype implementations of DisCo. The first prototype uses the exposure bracketing mode available in cameras for acquiring two exposures sequentially. This method, although straight-forward to implement on most digital cameras, is prone to errors due to inter-frame camera motion. Our second prototype is based on simultaneous dual exposure (SDE) sensors. SDE sensors have pixels with two different exposures interlaced with each other [Nayar and Mitsunaga 2000], and can *simultaneously* capture two exposures in a single image. These sensors are now also commercially available in consumer cameras [Toshiba 2013] for capturing high-dynamic-range images. This prototype acquires the signal as well as the display pattern from *a single image.*, and is thus robust to errors due to motion. This is illustrated in Figure 1. Since it uses easily available hardware for both sending and receiving messages, DisCo can be integrated into existing infrastructure, and readily adopted in several user-interfaces involving cameras and displays.

Scope and limitations: While designing a communication method, there is a trade-off between the data-rate and robustness. On one hand are methods based on high-speed photo-diodes [Elgala et al. 2009; Vucic et al. 2010] that can achieve high data rate, but only in controlled settings. On the other hand, in order to be applicable as a user interface, a communication method must be able to perform reliably in uncontrolled real-world situations, while potentially sacrificing the data-rate. This is shown in Figure 3. The focus of this paper is on designing user interfaces. DisCo can work robustly in challenging scenarios, such as when the display is significantly smaller/larger than the sensor field-of-view, occlusion, perspective distortion, camera rotation and de-

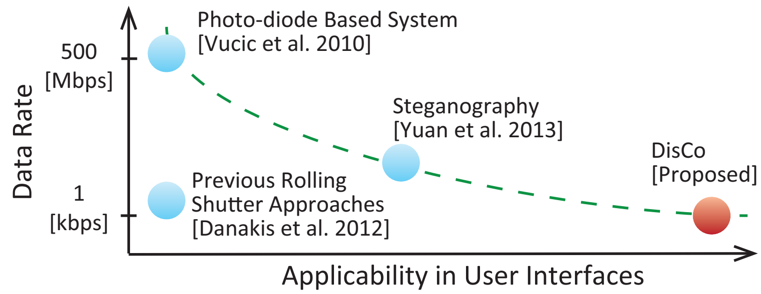


Fig. 3. **Tradeoff between Data Rate and Applicability in User Interfaces:** Communication techniques face a trade-off between their data-rate and robustness. Previous approaches based on photo-diodes can achieve high data rate, but in controlled settings such as indoor wireless networks. On the other hand, in order to be applicable as a user interface, a communication method must be able to perform reliably in uncontrolled real-world situations, while potentially sacrificing the data-rate. Since the focus of this paper is on designing user interfaces, the proposed method, DisCo, can work robustly in challenging scenarios, while requiring only low-cost consumer devices.

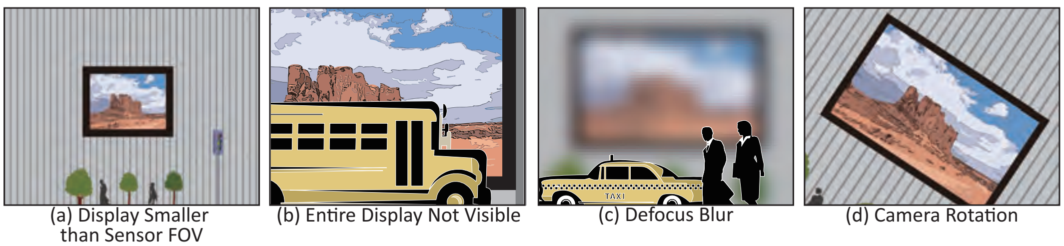


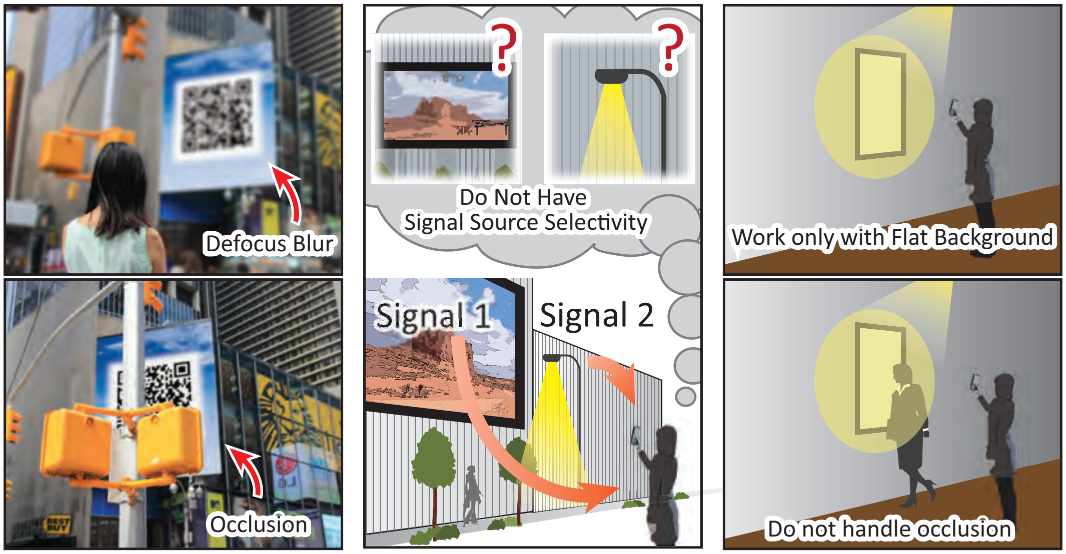
Fig. 4. **Display-Sensor Communication in the Wild:** DisCo performs robustly in challenging real-world situations such as when (a) the display is significantly smaller than the sensor field-of-view (FOV), (b) the display is partially visible to the sensor due to being outside the FOV, or due to occlusions, (c) there is camera defocus and (d) there is camera rotation.

focus (Figure 4). In addition to displays, any device emitting temporally modulated light such as a single or an array of LEDs (e.g., ceiling lights) can be used to convey information. DisCo can also operate in the 'spotlight' configuration where a reflective surface illuminated by a light source acts as the transmitter (Figure 2). The surface can have arbitrary shape and texture[1]. The data-rate achieved by DisCo is significantly lower than photo-diode based systems, but is sufficient to convey short messages such as URLs and pairing keys in a single image, which can enable several UI applications.

2. RELATED WORK

Spatial-domain steganography: One of the simplest techniques for embedding hidden information in displays is spatial-domain steganography (or watermarking), where a spatial code (e.g., QR-code [ISO 2006]) is embedded in the display image [Cheddad et al. 2010; Grundhofer et al. 2007; Yuan et al. 2013; Kamijo et al. 2008; Chan et al. 2010]. The performance of these techniques depends on the distance and inclination of the display with respect to the sensor [Perli et al. 2010]. Moreover, most systems require the entire display to be visible to the sensor. This constraint is a strong limitation since displays are often only partially visible to the sensor. These techniques

[1]One exception is if the scene is perfectly black. Since such a scene does not reflect any light, camera cannot receive the signal.



(a) Spatial Coding Methods (Steganography)
(b) Photo-diode Based Methods
(c) Previous Rolling Shutter Based Methods

Fig. 5. **Limitations of Previous Methods:** (a) Spatial-domain coding methods (e.g., steganography) cannot function reliably in the presence of defocus blur or occlusions. (b) Photo-diodes based systems require dedicated high-speed photo-diodes for communication. Since they do not capture images, they cannot differentiate between multiple signal sources in a scene. (c) Previous rolling shutter based methods are limited to work only with flat textureless displays, and are not robust to occlusions.

are also not robust to common imaging degradations, such as defocus blur (Figure 5 (a)).

Communication using temporally modulated light: In these systems, the signal transmitter is a light source modulated at high temporal frequencies. These methods require specialized sensors, such as photo-diodes [Elgala et al. 2009; Vucic et al. 2010] or high-speed cameras [Yoshimura et al. 2001; Matsushita et al. 2003; Kagawa et al. 2009; Sarkera et al. 2009]. Although photo-diodes can receive the transmitted signals, they cannot simultaneously capture images for human consumption. Moreover, since they do not capture images, photo-diode based systems cannot differentiate between multiple signal sources in a scene (Figure 5 (b)). High-speed cameras are expensive and cannot capture high spatial-resolution images. Hence, these techniques cannot be deployed in consumer devices such as cell-phones cameras. DisCo uses only low-cost off-the-shelf components, and can easily be incorporated into existing consumer imaging devices.

Rolling shutter sensors: Rolling shutter image sensors have recently been used for communicating with light sources [Nokia 2013; Woo et al. 2012; Danakis et al. 2012]. These methods, however, have limited applicability as they cannot work with general purpose displays which need to display a large range of images/text. This is because they assume that the transmitter has uniform brightness (spatially). Moreover, these assume that the light source occupies the entire sensor field-of-view, and thus cannot handle occlusion and small light sources(Figure 5 (c)). In contrast, DisCo is compatible with a significantly larger class of devices as transmitters, including displays and non-emitting surfaces with arbitrary texture.

Radio-wave based communication: Existing radio-wave based communication methods, such as Wi-Fi and Bluetooth achieve high data rate over long distances.

	Robustness to Blur, Occlusion and Display Geometry	Both Display and Light as Transmitters	Signal Source Selectivity
DisCo	✓	✓	✓
Spatial Coding Based Methods	✗	✗	✓
Photo-diode Based Methods	✗	✓	✗
Previous Rolling Shutter Based Methods	✗	✗	✓

Fig. 6. **Comparison with Previous Methods:** In this table, we compare various communication methods based on their robustness, the ability to use a wide range of light sources as signal transmitters, and the flexibility of selecting the signal source if multiple sources are present in the scene. Most previous communication approaches are optimized for achieving large data-rates, and are not robust to common imaging degradations. DisCo is designed to be robust, intuitive, and flexible, so that it can be widely applicable in user interfaces.

However radio-waves do not have directionality and can penetrate walls. In order to communicate using these modalities, the two devices must be "paired" by manually selecting the device and entering a password. This reduces the overall fluidity of the user-experience, which is critical in most consumer applications. Near-field-communication (NFC) methods perform pairing by bringing the devices close to each other. This physical requirement limits their applicability. With the proposed system, it would be possible to pair devices over large distances with a fast and intuitive 'point-and-pair' interface.

Comparison summary: The table in Figure 6 compares DisCo and several existing communication methods, based on their robustness, compatibility with different light sources as signal transmitters, and the flexibility of selecting the signal source if multiple sources are present. Most previous approaches are optimized for achieving large data-rates, and are not robust to common imaging degradations. Moreover, previous approaches are compatible with only a small set of sources as transmitters (only displays or only uniform background light sources) and receivers (e.g., only high speed sensors). In comparison, DisCo is designed to be robust and compatible with a wide range of sources, and thus ideally suited for user interface applications.

3. IMAGE FORMATION MODEL

DisCo consists of a spatio-temporally modulated display (transmitter) and a rolling shutter sensor (receiver). The display brightness is temporally modulated with the function $f(t)$, which encodes the signal to be transmitted. We call $f(t)$ the signal function. Conceptually, the display can be thought of as having two layers - a signal layer and a texture layer. The texture layer consists of the image that is displayed to humans. This is illustrated in Figure 7 (a). The display could be realized either as an LCD panel with a temporally modulated LED back-light, or as a single LED, or even with a spotlight shining on a reflective surface, e.g., a painting on a wall (Figure **??**(a)). In the last case, the illuminated part of the surface is considered the display and the texture of the surface forms the texture layer.

In the following, we assume that the display completely occupies the sensor field-of-view so that every sensor pixel receives light only from the display area. This assumption is made only for ease of exposition, and is not a requirement of the proposed

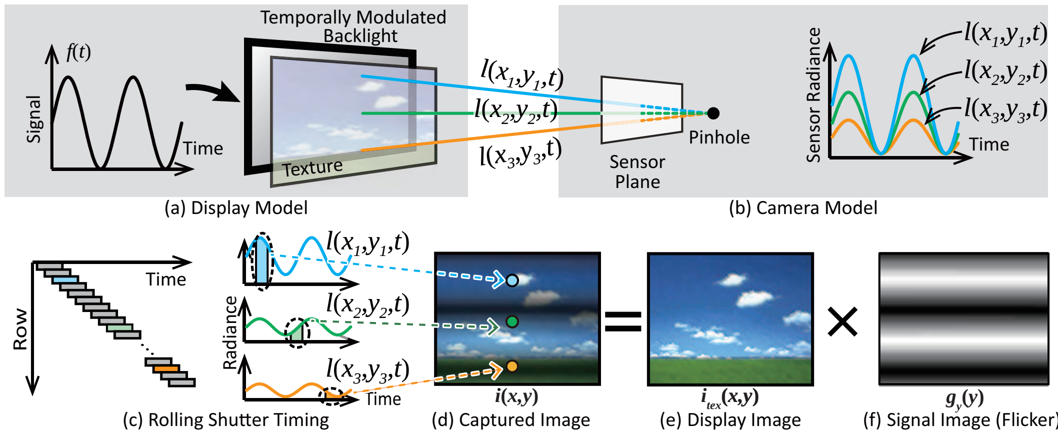


Fig. 7. **System Overview:** DisCo consists of a spatio-temporally modulated display (transmitter) and a rolling shutter sensor (receiver). (a) The display is modeled as having two layers - the texture layer which is an image to be viewed by humans, and the backlight layer that conveys the signal. The intensity of the backlight is temporally modulated with the signal $f(t)$. The temporal frequency of $f(t)$ is significantly higher than what humans can perceive, and thus they only see the image displayed by the texture layer. (b) $l(x, y, t)$ is the temporal radiance profile incident at camera pixel (x, y) at time t. Because the entire display is modulated by the same temporal function, the radiance profiles for different pixels differ only by a multiplicative scale factor, which is determined by the display pattern. (c) Due to the rolling shutter, pixels in different rows sample the temporal radiance profiles at different instants. (d) This creates a spatial flicker in the captured image. We show that the captured image can be factorized as a product of the flicker-free display image (e) and the flicker image (f). The flicker image contains the information embedded in the temporal signal $f(t)$.

method [2]. Let $l(x, y, t)$ be the radiance incident at sensor pixel (x, y), at time t. [3] This is illustrated in Figure 7(a). Because the entire display is modulated by a single temporal function $f(t)$, the radiance $l(x, y, t)$ can be factorized into spatial and temporal components:

$$l(x, y, t) = l_{tex}(x, y)\, f(t)\,, \tag{1}$$

where $l_{tex}(x, y)$ is the amplitude of the temporal radiance profile at pixel (x, y), and is determined by the display's texture layer. Note that the temporal radiance profiles at different sensor pixels differ only in their amplitudes $l_{tex}(x, y)$. This is illustrated in Figure 7 (b).

Let $e(x, y, t)$ be the exposure function at pixel (x, y). If pixel (x, y) is *on* (i.e., it captures light) at time t, $e(x, y, t) = 1$, otherwise, if the pixel is *off* (i.e., blocks incident light) $e(x, y, t) = 0$. The measured brightness value $i\,(x, y)$ is:

$$i\,(x, y) = k \int_{-\infty}^{\infty} l\,(x, y, t)\, e(x, y, t) dt\,, \tag{2}$$

where k is the sensor gain that converts radiance to pixel brightness. Since the sensor has a rolling shutter, different rows capture light during different, shifted time inter-

[2]In general, a sensor pixel may receive light from outside the display, due to the display not completely occupying the sensor's FOV or due to occlusions. It can be shown that the image formation model for the general case has the same form as that of the special case where pixels receive light only from the display. For details, the reader is referred to the supplementary technical report.

[3]For simplicity, a single color channel is considered. For colored sensors, similar analysis can be done individually for each color channel.

vals. The amount of shift is determined by the row index y and the speed of the rolling shutter. The exposure function $e(x, y, t)$ can be modeled as a time-shifted function $s(t)$:

$$e(x, y, t) = s(t - t_y)\,, \tag{3}$$

where t_y is the temporal shift for a pixel in row y. The exposure timing of a rolling shutter sensor is illustrated in Figure 7 (c). The function $s(t)$, called the shutter function, can be a rect function, a temporal Gaussian, or even a high frequency binary code [**?**]. Substituting Eqs. 1 and 3 in Eq. 2, we get:

$$\begin{aligned} i\,(x, y) &= k \;\; l_{tex}\,(x, y) \int_{-\infty}^{\infty} s\,(t - t_y)\, f\,(t)\, dt \\ &= k \;\; l_{tex}\,(x, y) \;\; g'(t_y)\,, \end{aligned} \tag{4}$$

where $g'(t_y) = (s * f)\,(t_y)$ is the convolution of the signal and the shutter functions. $g'(t_y)$ is a function of the temporal shift t_y, which in turn depends on the sensor row index y. Typically, $t_y = \frac{y}{r}$, where r rows/second is the speed of the rolling shutter. We rewrite the above equation as:

$$i\,(x, y) = \underbrace{i_{tex}(x, y)}_{\text{display image}} \times \underbrace{g(y)}_{\text{signal image}}\,, \tag{5}$$

where $i_{tex}(x, y) = k \times l_{tex}(x, y)$ is called the display image as it is determined by the image being displayed, and $g(y) = g'(t_y) = (s * f)\,(t_y)$ is the signal image that encodes the signal function $f(t)$. The above equation states that the texture and the signal layers of the display are observed as two separable (and unknown) components: the display image and the signal image. This is illustrated in Figures 7 (e)-(f). The temporal signal $f(t)$ manifests only in the signal image $g(y)$, and the display's texture layer is captured only in the display image $i_{tex}(x, y)$. Eq. 5 is the image formation model for DisCo and forms the basis of our method.

Structure of the signal image: The signal image $g(y)$ varies only along the y dimension because different sensor rows sample the signal function $f(t)$ at different instants (Figure 7 (d)), and thus have different intensities. However, all the pixels in a given row sample $f(t)$ at the same instant, and thus have the same intensity. As a result, $g(y)$ has the form of a horizontal *flicker* pattern, as illustrated in Figure 7 (f).

Since the signal image $g(y)$ is one-dimensional (1-D), for computational efficiency, we perform analysis on horizontal sum images which are 1-D signals, i.e., $i(y) = \sum_x i(x, y)$ and $i_{tex}(y) = \sum_x i_{tex}(x, y)$. Saturated image pixels are excluded from the summation. Then, Eq. 5 can be written as $i(y) = i_{tex}(y) \times g(y)$. For the rest of the paper, we use this 1-D form of the image formation equation.

Invariance of the signal image to display-camera geometry, partial occlusions, and imaging parameters: The image formation model in Eq. 5 is derived without making any assumptions about the display's shape, orientation or location with respect to the sensor, or about imaging parameters such as zoom and defocus. Since the signal component $g(y)$ depends only on the signal function $f(t)$ and the shutter function $s(t)$, any changes in display-sensor geometry or imaging parameters (zoom and focus) manifest only in the display image $i_{tex}(x, y)$. Specifically, display's orientation and location determine the shape of display's projection in the captured image, sensor zoom influences the size of the display's projection, and camera focus determines the amount of blur in the display image [4]. If the display is partially occluded so that it is visible to a (non-empty) subset of pixels in each sensor row, because the

[4]It can be shown that the signal image $g(y)$ is invariant to camera defocus. For a derivation, see Appendix.

(a) Large Camera Distance (b) Camera Rotation (c) Camera Zoom / Defocus

Fig. 8. **Invariance of Signal (Flicker) Image to Imaging Geometry:** The flicker pattern in the signal image is invariant to imaging geometry parameters, such as (a) display-camera distance, (b) camera rotation, (c) camera zoom and defocus blur. In these examples, the display back light was temporally modulated with a 500 Hz. sinusoid. This results in a spatial sinusoidal flicker. Notice that the period of the flicker, h_{sine}, is the same in all the cases.

captured image is summed horizontally, the signal image $g(y)$ is still sampled at every row location. If $\alpha_y > 0$ is the fraction of pixels in sensor row y that see the display, the amplitude of the signal image is scaled by α_y. Under mild assumptions, α_y can be assumed to be locally constant, and absorbed in the display image (see supplementary technical report for details).

As a result, the signal image is always a horizontal flicker pattern. *Its functional form and structure are invariant to the display-camera geometry, partial occlusions and camera parameters.* A few examples are illustrated in Figure 8. In the shown examples, $f(t)$ is a 500 Hz. sinusoidal signal and the shutter $s(t)$ is a rect function of width 0.5 ms such that $s(t) = 1$ when $0 \leq t \leq 0.5$ ms, otherwise $s(t) = 0$. This results in a sinusoidal flicker pattern. Notice that the period of the flicker, h_{sine}, is independent of camera-display geometry or camera zoom. Even if only a small fraction of the display is visible to the camera due to large zoom (Figure 8(c)), the flicker image retains the same structure, and captures the information contained in the signal function.

4. SIGNAL RECOVERY BY CAPTURING TWO DIFFERENT EXPOSURES

In order to decode the information in the signal image $g(y)$, we need to separate it from the display image $i_{tex}(y)$. Since both signal and display components are unknown, in general, they cannot be separated from a single captured image. The key idea is that if we capture two images $i_1(y)$ and $i_2(y)$ with two different shutter functions $s_1(t)$ and $s_2(t)$, we can get two different equations, which enable performing the separation. The two images can be captured sequentially using the exposure bracketing mode available in most digital cameras. This approach, while suitable for static scenes and cameras, is prone to errors if there is scene/camera motion. As we will describe later in Section 5, in order to deal with motion, we propose using a camera that captures two images with different exposure functions *simultaneously in a single shot*.

The two images are given as:

$$i_1(y) = i_{tex}(y) \times (s_1 * f)(t_y) \quad (6)$$

$$i_2(y) = i_{tex}(y) \times (s_2 * f)(t_y) \quad (7)$$

This is a system of two equations in two unknowns: signal $f(t)$ and the flicker-free display image $i_{tex}(y)$. Since the shutter functions $s_1(t)$ and $s_2(t)$ are known, these two equations can be solved simultaneously to recover both $f(t)$ and the flicker-free image $i_{tex}(x, y)$. In the following, we provide details of the signal recovery algorithm.

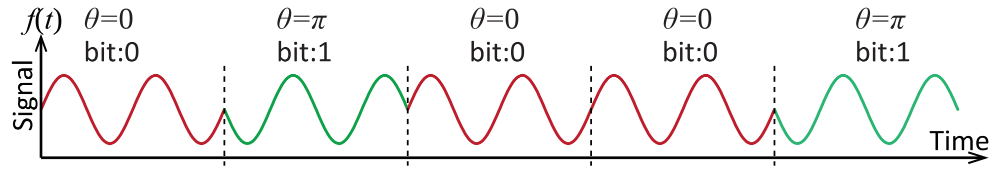


Fig. 9. **Signal Coding Method:** We use the phase-shift keying (PSK) signal coding technique, where information is embedded in the phase of sinusoidal signals. For example, in binary PSK, the phase θ of sinusoids takes binary values (0 and π), thus encoding binary bits. Bits are transmitted sequentially in time.

4.1. Signal Model and Recovery Algorithm

We consider the signal $f(t)$ to be a sum of sinusoids of different frequencies (the set of frequencies is typically a small, discrete set). This signal encoding scheme is called orthogonal-frequency-division-multiplexing (OFDM) [Nee and Prasad 2000], and is one of the most popular schemes in communication literature [5]. For each frequency, information is embedded in the phase of the sinusoids. This method of embedding information is called *phase-shift keying*. For instance, in binary phase-shift keying, binary symbols of 0 and 1 are embedded by using sinusoids of phase 0 and π, respectively. Bits (sinusoids with different phases) are transmitted sequentially in time. An example for a single frequency is illustrated in Figure 9.

Since we use sinusoidal signals, for computational efficiency, we perform computations in the Fourier domain. Eqs. 6-7 can be written in the Fourier domain as:

$$I_1(\omega) = I_{tex}(\omega) * (S_1(\omega) F(\omega)) \quad (8)$$

$$I_2(\omega) = I_{tex}(\omega) * (S_2(\omega) F(\omega)) \;, \quad (9)$$

where ω is the spatial frequency. The functions denoted by uppercase letters are the Fourier transforms of the functions denoted by the corresponding lower case letters. These two equations can be combined as:

$$I_1(\omega) * (S_2(\omega) F(\omega)) - I_2(\omega) * (S_1(\omega) F(\omega)) = 0 \,. \quad (10)$$

The temporal signal $f(t)$ consists of a small, discrete set of temporal frequencies $\Omega = [\omega_1, \ldots, \omega_M]$. We need to solve Eq. 10 only for the frequency set Ω. Let $\vec{I_1}$ be the vector of values $[I_1(\omega_1), \ldots, I_1(\omega_M)]$. The vectors $\vec{I_2}$, $\vec{S_1}$, $\vec{S_2}$ and $\vec{F}$ are defined similarly. By observing that convolution can be expressed as multiplication by a Toeplitz matrix and element-wise multiplication as multiplication by a diagonal matrix, Eq. 10 can be compactly represented in matrix form as:

$$(\mathbf{I}_1\mathbf{S}_2 - \mathbf{I}_2\mathbf{S}_1)\,\vec{F} \;=\; 0 \,, \quad (11)$$

where $\mathbf{I}_1$ and $\mathbf{I}_2$ are Toeplitz matrices defined by vectors $\vec{I_1}$ and $\vec{I_2}$, respectively. $\mathbf{S}_1$ and $\mathbf{S}_2$ are diagonal matrices defined by vectors $\vec{S_1}$ and $\vec{S_2}$, respectively.

The matrices $\mathbf{I}_1$ and $\mathbf{I}_2$ are defined by captured image intensities and $\mathbf{S}_1$ and $\mathbf{S}_2$ are defined in terms of the known shutter functions. The goal is to recover the unknown vector $\vec{F}$. The above equation can be solved as a linear system of the form $\mathbf{AX} = 0$. In order to avoid the degenerate solution ($\vec{F} = 0$) and ambiguity (if $\vec{F}$ is a solution, then $s\vec{F}$ is also a solution for any complex number s), we impose the constraint that $F(0) = 1.0$, i.e., the DC level of the signal $f(t)$ is 1.0.

Recall that the signal comprises of multiple bits that are transmitted sequentially, and are thus captured at different spatial locations in the signal image. We recover

[5] In general, the proposed system can work with any of the several signal encoding schemes proposed in the communications literature.

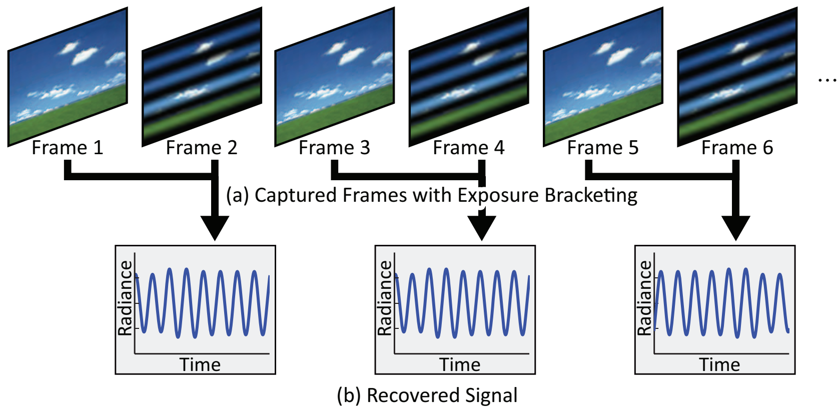


Fig. 10. **Input and Output of DisCo with Exposure Bracketing Mode:** (a) DisCo can be implemented by using the exposure bracketing mode available in most existing digital cameras. two images with different exposures (one long, and one short) are captured sequentially. The long exposure is chosen so that the captured image is nearly flicker-free (Frames $1, 3, 5, \ldots$). (b) The two images are then divided to recover the signal image, from which the temporal signal is estimated.

each bit individually by applying the signal recovery algorithm to a small interval of the captured image at a time. The interval-size h_{bit} is the number of image rows required to encode a single bit. h_{bit} is determined by the signal frequency; higher the frequency of $g(y)$ (due to $f(t)$ having high temporal frequency), smaller the interval-size. Thus, we divide the captured images $i_1(y)$ and $i_2(y)$ into small 1-D intervals, and recover $\vec{F}$ by computing Eq. 11 on each interval individually. Since computations are done locally, $I_1(\omega)$ and $I_2(\omega)$ are the short time Fourier transforms (STFT) of $i_1(y)$ and $i_2(y)$. Once $\vec{F}$ is computed, we recover the signal $f(t)$ and the embedded information by applying inverse Fourier transform. The display image $i_{tex}(x, y)$ is then computed by using Eq. 5: $i_{tex}(x, y) = \frac{i(x,y)}{g(y)} = \frac{i(x,y)}{(s*f)(t_y)}$.[6]

Simulations: We evaluated the performance of the signal recovery algorithm using several simulations. The simulations illustrate that brighter and larger display images result in high SNR communication between the display and the sensor. Also, although higher signal frequencies may achieve higher data rate, they result in lower SNR. This tradeoff must be considered while designing practical systems. **For detail, please see the supplementary technical report.**

4.2. Capturing Two Exposures With Exposure Bracketing

Most existing digital cameras have an exposure bracketing mode for capturing high-dynamic-range (HDR) images, where multiple images with different exposures are captured sequentially. We use the exposure bracketing functionality for capturing the two different exposures required for DisCo. However, because the two images are taken sequentially, the second image samples the emitted temporal signal at a different time instant than the first image, and thus captures a different temporal signal $f'(t)$. The two images are given as:

$$i_1(y) = i_{tex}(y) \times (s_1 * f)(t_y) \quad (12)$$
$$i_2(y) = i_{tex}(y) \times (s_2 * f')(t_y) \quad (13)$$

[6]If one of the shutter functions is significantly longer than the period of the signal $f(t)$, the corresponding $g(y)$ will be approximately constant. In that case, the corresponding captured image $i(x, y)$ is nearly flicker-free, and can directly be used as the display image.

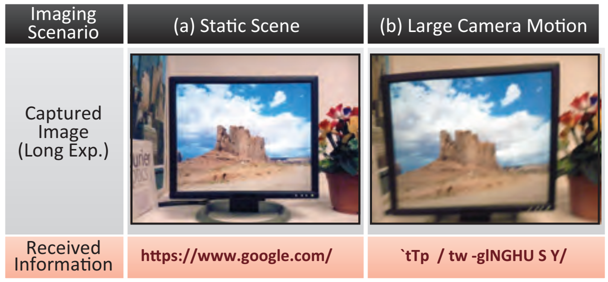


Fig. 11. **Experimental Demonstration of DisCo with Exposure Bracketing:** The two exposures are 0.25 ms and 16 ms. The message *https://www.google.com/* is embedded in the phases of sinusoids of frequencies 1 and 2 kHz. (a) If the scene and camera are static (or if the motion is small), the signal is recovered accurately. (b) However, if there is strong camera motion, the images cannot be registered reliably, resulting in incorrect signal recovery.

Since our decoding algorithm assumes that both the images observe the same signal $f(t)$, it cannot recover the signal. This problem is solved by capturing two images i_{short} and i_{long} with alternating short and long exposures, s_{short} and s_{long}, respectively, as shown in Figure 10. If s_{long} is chosen so that it is significantly longer than the period of the temporal signal, the signal image $g_{long}(y) = (s_{long} * f)$ is approximately constant, irrespective of the time instance when the signal is sampled. Thus:

$$(s_{long} * f)\,(t_y) \approx (s_{long} * f')\,(t_y) \approx K\,, \tag{14}$$

where K is a constant. By using the above approximation, the two images i_{short} and i_{long} can be expressed as:

$$i_{short}\,(y) \;=\; i_{tex}\,(y) \times (s_{short} * f)\,(t_y) \tag{15}$$

$$\begin{aligned} i_{long}\,(y) \;&=\; i_{tex}\,(y) \times (s_{long} * f')\,(t_y) \\ &\approx\; i_{tex}\,(y) \times (s_{long} * f)\,(t_y) \end{aligned} \tag{16}$$

Eqs. 15 and 16 are the same as in Eqs. 6 and 7. Thus the signal $f(t)$ can be estimated using the same algorithm as given in Section 4.1. Note that the data transmit rate is halved since only the signal transmitted during the capture of short exposure frames is decoded. [7]

Scene and camera motion: The implementation using exposure bracketing assumes that both the scene and the camera are static while the two images are captured. If there is scene/camera motion during capture, the images need to be aligned by computing relative motion between them. Unfortunately, if the inter-frame motion is large, image alignment techniques often produce inaccurate results. This can result in erroneous signal recovery. Figure 11 shows an example using a conventional camera with exposure bracketing mode. The two exposures are 0.25 ms and 16 ms. The message *https://www.google.com/* is embedded in the phases of sinusoids of frequencies 1 and 2 kHz. If the scene and camera are static, the exposure bracketing based implementation recovers the signal accurately. However, if there is strong camera motion (e.g., due to hand shake), the images cannot be registered reliably, resulting in incorrect signal recovery.

[7] Because $i_{long}(x, y)$ can be approximated as the texture image, it is also possible to estimate flicker component by calculating image ratio $i_{ratio}(x, y) = \frac{i_{short}(x,y)}{i_{long}(x,y)} \approx \frac{g_{short}(y)}{K}$.

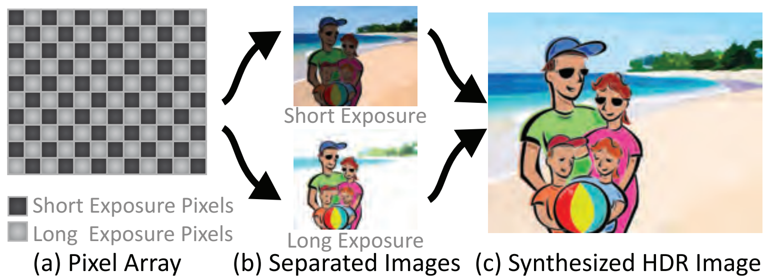


Fig. 12. **Simultaneous Dual Exposure (SDE) Sensor:** (a) SDE sensors have an array of pixels with two different exposures. (b) This allows them to capture images with different exposures in a single-shot, and then synthesize a high-dynamic-range (HDR) image (c). Such sensors are now available in consumer cameras due to their iability for capturing HDR images.

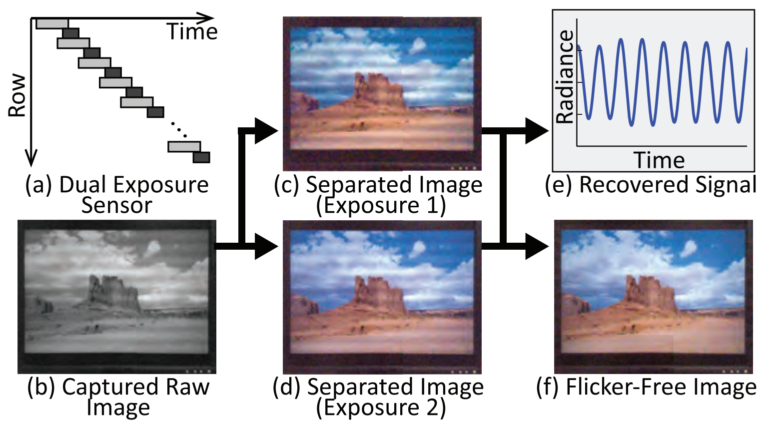
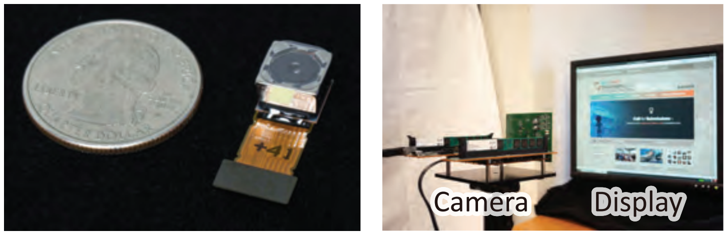


Fig. 13. **Input and Output of DisCo with SDE:** (a) Exposure timing of an SDE rolling shutter sensor with two different exposures. (b) An example captured image (input) with a rolling shutter SDE sensor. The display was modulated with a 500 Hz. temporal sinusoid. (c-d) Images of different exposures are extracted from (b), from which the signal (e) and the flicker-free image (f) are recovered (output).

5. CAPTURING TWO EXPOSURES WITH SIMULTANEOUS DUAL EXPOSURE SENSOR

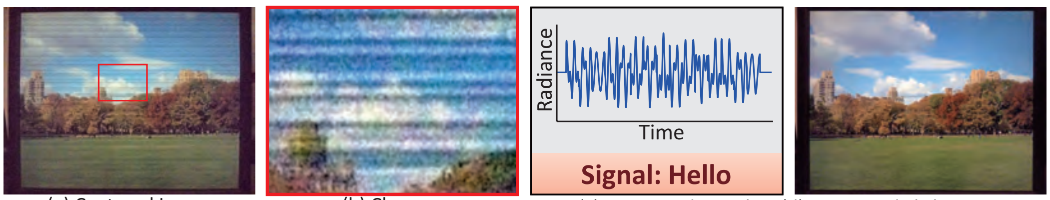
In order to avoid errors in the recovered signal due to motion, the two images with different exposures must be captured simultaneously. One way to achieve this is by using two synchronized cameras that are co-located using additional optics. Although theoretically feasible, this is not a practically viable option, especially in consumer settings.

We propose capturing two different exposures in *a single image* by using a simultaneous dual exposure (SDE) sensor. An SDE sensor has an array of pixels with two different exposures interlaced with each other. Such sensors are now commercially available [Toshiba 2013], and are expected to be deployed in future consumer cameras given their ability to capture high dynamic range images in a single-shot. An example SDE sensor with 2 different exposures (long and short) is shown in Figure 12. We use an SDE sensor for capturing two sub-images with different exposures in a single image. Figure 13 shows an SDE image captured with our prototype. Notice the flicker in the two extracted sub-images. Using the two sub-images as input, the recovery algorithm estimates the transmitted signal and the flicker-free display image.



(a) SDE Image Sensor Module (b) Prototype of DisCo

Fig. 14. **Hardware Prototype:** (a) The SDE image sensor module. The sensor can capture two different exposures in a single image. (b) A prototype of DisCo with the display and the camera. We have also built a prototype with a single LED as source, and a prototype in the spotlight configuration for non line-of-sight communication (Figure 4 (f)). Please see the supplementary video for results.



(a) Captured Image (b) Close-up (c) Recovered Signal (d) Recovered Flicker-Free Image

Fig. 15. **Experimental Demonstration of Single Image Communication:** (a) An image captured using our prototype. (b) Close-up of (a). Notice the flicker pattern. (c) Recovered temporal signal and the embedded text "Hello". (d) Recovered flicker-free display image.

5.1. Implementation Details

In this section, we will describe a prototype for DisCo using a temporally modulated LCD display and an SDE sensor with rolling shutter (Figure 14). If readers do not interested in the implementation, please skip to Sec. 5.2. The temporally modulated display was implemented by replacing the back-light of a Dell LCD monitor by an LED array. The LED array is driven with an RECOM RCD-24-1.2 LED driver. We have also developed a prototype where the signal source is a single LED, as well as a prototype in the spotlight configuration where the sensor receives light after reflection from a scene. The results for these prototypes are shown in **the supplementary video.**

In all the prototypes, the sensor has a rolling shutter. It acquires a single image within which two sub-images of different exposures are interleaved. The two sub-images are separated in a post-processing step from the raw captured image.

Temporal synchronization: If the sensor and the display are not temporally synchronized, the start of the transmitted signal cannot be localized in the signal image, and the signal cannot be decoded. In order to handle lack of synchronization, we use two well-known techniques in communications literature. First, a *pilot symbol* is embedded in the signal to determine the beginning of the signal. In our implementation, the pilot symbol is a sinusoid of a frequency that is not used to encode the main signal, so that it is readily detected. Second, *guard interval* based synchronization [van De Beek et al. 1997] is used to determine the start of every symbol (bit). In this scheme, the end of each symbol is copied to its beginning. Then, by self-correlating the signal with itself, the beginning location of every symbol is computed.

Error detection: There are several sources of errors in the signal recovery process (Eq. 11), namely, sensor saturation, low display brightness, small display area and sensor noise. Moreover, although the recovery algorithm is robust to partial occlusions, severe occlusions where none of the pixels in a sensor row sees the display can lead

Imaging Scenario	(a) Static Scene	(b) Large Camera Motion
Captured Image		
Received Information	https://www.google.com/	https://www.google.com/
Number of Frames	27 [frames]	35 [frames]
Data Rate	237 [bps]	183 [bps]

Fig. 16. **Performance of SDE Sensor-Based DisCo in the Presence of Camera Motion:** The camera motion is the same as in Figure 11. While exposure bracketing based implementation can recover the signal reliably only when the camera is static, the SDE based implementation achieves accurate results even in the presence of large camera motion. Note that more frames are required for signal recovery when there is motion, as compared to when the scene is static.

to errors. Finally, if the display occupies only a small area in the captured image, the signal image has low amplitude, and the recovered signal has low SNR. In all these scenarios, the recovered signal may have errors, which must be detected. Let the recovered solution for a region of the captured image be $\vec{F}$. We detect errors by computing the left hand side of Eq. 11, $\left((\mathbf{I}_1 \mathbf{S}_2 - \mathbf{I}_2 \mathbf{S}_1) \vec{F} \right)$. If the value is greater than a prescribed threshold, we declare the recovered signal to be erroneous.

Achieving robustness to occlusion: DisCo handles occlusions by creating redundancy in the transmitted signals. The display transmits the signal $f(t)$ repeatedly and the sensor capturing a sequence of frames (assuming small inter-frame motion). The signal length is optimized so that a particular signal bit is captured in different image locations in successive captured images. Since the errors are location specific (due to occlusions or low texture brightness), if a bit is decoded incorrectly in one frame, with high probability, it will be decoded correctly in a subsequent frame. The number of frames that need to be captured depends on the signal size, the extent of occlusions, the brightness of background texture, display area and sensor noise. If the signal size is relatively short and the occlusion is only partial, a single image is sufficient. For more challenging situations (e.g., large signal size, strong occlusions, small display areas or strong sensor noise), multiple images are used. Even for such situations, DisCo degrades the data rate gracefully, instead of entirely preventing communication. Examples demonstrating the functioning of DisCo in various levels of occlusion are shown in the results section.

5.2. Results

Communication using a single captured image: Figure 15 shows results of display-sensor communication using a single captured image with the SDE sensor. The display sends the message *Hello*, and the sensor captures a single SDE image with two interleaved exposures (0.25 ms and 16 ms). By applying the signal recovery algorithm, the signal and flicker-free scene image are separated.

Signal recovery in the presence of motion: Figure 16 shows the performance of SDE sensor based DisCo when the camera is moving. The camera motion is the same as in Figure 11. While exposure bracketing based implementation can recover the signal reliably only when the camera is static, the SDE based implementation achieves

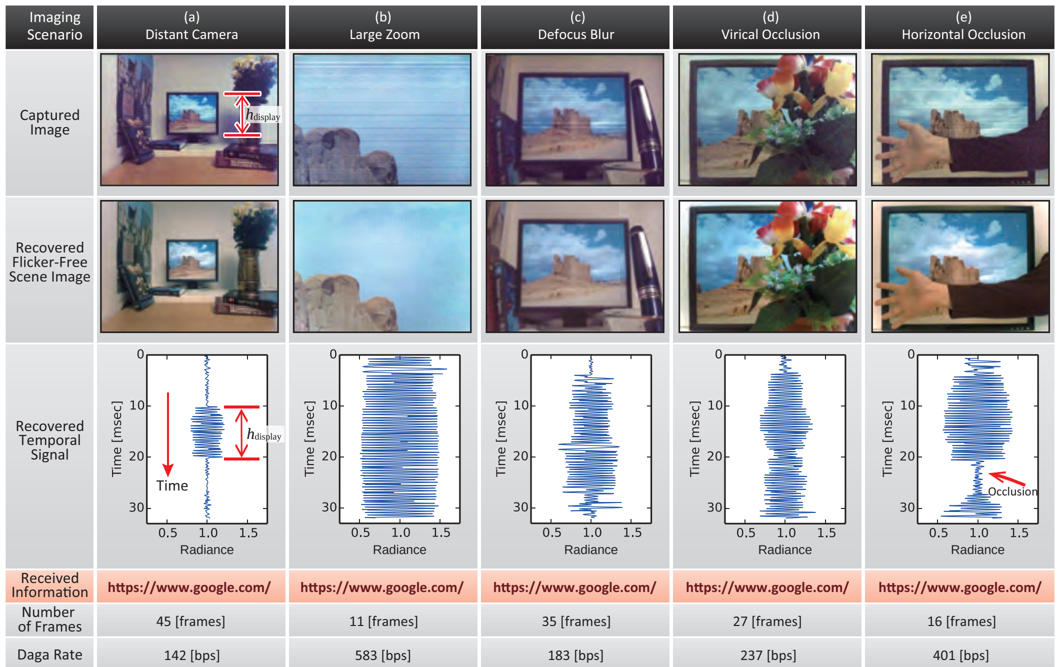


Fig. 17. **Experimental Demonstration of DisCo in Challenging Imaging Scenarios:** We have extensively tested the proposed system in a variety of real-world imaging situations. (a) Display smaller than camera's FOV due to large display-camera distance. (b) Display larger than camera's FOV due to large zoom/small display-camera distance. (c) Display blurred due to camera defocus. (d-e) Occluding objects between camera and display. In these examples, the signal (URL https://www.google.com/) was transmitted and received by the camera. The system adapts to challenging conditions by capturing multiple frames. In order to capture the high frequency temporal signal, one of the two exposures is short. As a result, the captured image appears noisy. However, since the other exposure is long, the recovered flicker free image has significantly lower noise. **For more results, please see the supplementary video.**

accurate results even in the presence of large camera motion. Note that as compared to static scene, more frames are required for signal recovery when there is motion.

Performance "in the wild": We have evaluated our system in different challenging real-world situations. Figure 17 shows several examples. In each case, the display sent the URL "https://www.google.com/" and the sensor received it without error. A single modulation frequency of 2 kHz was used. The frequency of the pilot symbol (used for synchronization) was 1 kHz. The number of required frames depends on the display area and the shape of occlusions. In every example, the URL was received within 1.5 seconds (the sensor frame rate is 30 fps). **For more results, see the supplementary video.**

Communicating with a light source and a reflective surface: We have also developed prototypes where a single LED, as well as a non-emitting reflective surface serve as the signal transmitters, as illustrated in Figure 4 (e) and (f), respectively. Figure 18 shows experimental results for these two cases. When the light source is bright so that the camera observes a high SNR image, it is possible to use multiple frequencies that enable a higher data-rate. Figure 18 (a) and (b) used four ($[1, 2, 3, 4]$ kHz) and two ($[1, 2]$ kHz) frequencies for embedding information, respectively. In the spotlight configuration (Figure 18 (b)), an LED lamp illuminating a photograph on the wall was used to tag the photograph with meta-information (the time and location of

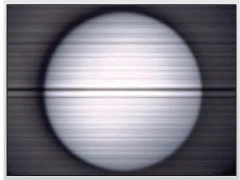
Imaging Scenario	(a) Communicating with Light Source	(b) Non Line-of-Sight Communication
Captured Image (Close-up)		
Received Information	https://www.google.com/	Fall 2014, Central Park, New York
Frequencies	1, 2, 3, 4 [kHz]	1, 2 [kHz]
Number of Frames	4 [frames]	7 [frames]
Daga Rate	1.6 [kbps]	1.1 [kbps]

Fig. 18. **Communicating with a light source and a reflective surface:** DisCo can use a single LED, as well as a non-emitting reflective surface as the signal transmitters, as illustrated in Figure 4 (e) and (f), respectively. (a) When directly communicating with a bright light source, because of high SNR, high data-rate is achieved. (b) In the spotlight configuration, an LED lamp shining on a photograph is used to tag it with meta-information (the time and location of the photograph). The information was received by the camera viewing the photograph. Since the illuminated surface can have arbitrary shape and texture, this functionality can be used in museums and stores for tagging objects.

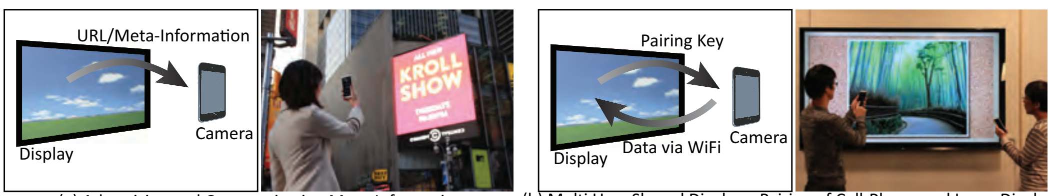


Fig. 19. **Application Scenarios:** (a) DisCo can be used for marketing and conveying meta-information (e.g., URLs, schedules of shows, item-prices and availability) via LED-based billboards installed in public places. (b) It can also be used for pairing of cell-phones screens with displays. Such pairing can allow a user to have a large display as an extension of their small cell-phone screens, or multiple users to share a large display among them for collaborative use.

the photograph). The information was received by the camera viewing the photograph.

6. APPLICATION SCENARIOS

DisCo can be used in a wide range of applications. **Please see the supplementary video for illustrations.** Following are a few examples.

Advertising and communicating meta-information: DisCo can be used for marketing and conveying meta-information (e.g., URLs, schedules of shows, item-prices and availability) via LED-based billboards installed in public places. This is illustrated in Figure 19 (a). Users can receive the information by simply pointing their cell phone towards the display. Currently, spatial codes (e.g., QR-codes) are used for these purposes. But DisCo does not require a dedicated space for displaying the QR-code, thus saving physical space and avoiding the display of unattractive patterns.

Pairing of devices: DisCo can be used for pairing of cell-phones screens with displays. This can allow a user to have a large display as an extension of their small cell-phone screen, and also to share a large display with other users [8]. Pairing of cell-phone devices with a display can be achieved by the display broadcasting its unique pairing key. One (or more) users can receive the key simply by pointing their phones towards the display. Once the key is received, pairing is established, and cell-phones can send data (e.g., images, videos) to be displayed on the screen using an existing communication modality such as Wi-Fi or blue-tooth. Figure 19(b) shows an illustration. In current popular techniques, the pairing key needs to be manually entered (e.g., Wi-Fi passwords), which makes the user experience less fluid. QR code also can show the passwords using display, however it limits the design of display image. The proposed technique provides a fast and intuitive interface for pairing of different devices.

Interactive Presentation: Another application can be during a presentation being given on a large screen. Current presentation scene allows a presenter to use the screen, and the audience hardly share the information using the screen. Using DisCo, a member of the audience can pair their cell-phone/laptop with the screen and show relevant information (e.g., additional charts, web-pages). Since light cannot penetrate walls, only the audience members present in the meeting room will have access to the screen.

Museums: DisCo is capable of non line-of-sight communication where a spotlight shining on a surface conveys meta-information about it. The information can be received by a user by simply pointing their cell-phones at the surface. This functionality can be used in museum settings; strategically installed spotlights can serve the dual purpose of enhancing the artifacts' appearance, while simultaneously communicating information about them in an unobtrusive manner.

Indoor navigation and location specific services: DisCo can utilize single or arrays of light sources (e.g., ceiling lights) as transmitters. The light sources, in addition to providing illumination, also broadcast their location (or other location specific information). Thus, it can be used for indoor navigation where GPS is known to be unreliable.

7. DISCUSSION AND LIMITATIONS

We have proposed DisCo, a novel display-sensor communication technique. It works with commercially available image sensors (both conventional and SDE) and a variety of sources (display, lights, illuminated surfaces), and performs reliably in difficult real-world scenarios. We have built hardware prototypes and shown several potential applications. Since DisCo observes image as well as signal to the users and it requires only off-the-shelf components, we believe it can be widely adopted in various user-interfaces involving sensors and displays. In the following, we discuss some limitations of DisCo.

Modulation frequencies and signal length: The sensor must know the modulation frequencies of the light sources a priori. This can be achieved by establishing communication standards in display-sensor communication, so that a fixed set of modulation frequencies are used. This is similar to radio-wave based communication modalities (e.g., Wi-Fi) where the modulation frequencies are pre-specified. It may be possible to relax this restriction and use arbitrary modulation frequencies. For this, the sensors must be able to compute the modulation frequency before performing decoding. While this functionality will provide more flexibility and will increase the available bandwidth, it may come at the cost of decreased data rate.

[8] If there are multiple external displays available, a user can select the desired display by selecting it in the captured image.

Currently, we also assume that the signal length is known to the receiving sensor. The receiver may be able to estimate the signal length automatically by measuring the time difference of arrival between two consecutive pilot symbols. This is an interesting avenue for future work.

Number of displays: So far, we have assumed that the sensor communicates with a single source (display) at a time. This limits the data transfer rate. It is possible to communicate via multiple sources simultaneously and thus achieve higher data rates by segmenting the captured image into different display regions.

Shutter function and SNR: We have limited ourselves to rect (pill-box) shutter functions due to ease of implementation. However, the algorithm is not limited to pill-box functions. It is possible to achieve higher SNR by using temporally coded shutter functions [**?**]. Also, we used a pair of exposures computed using a search based procedure, which may not be optimal. Designing theoretically optimal shutter functions requires further analysis of the image formation process, and is another promising future work direction.

REFERENCES

Li-Wei Chan, Hsiang-Tao Wu, Hui-Shan Kao, Ju-Chun Ko, Home-Ru Lin, Mike Y. Chen, Jane Hsu, and Yi-Ping Hung. 2010. Enabling Beyond-surface Interactions for Interactive Surface with an Invisible Projection. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 263–272. DOI:http://dx.doi.org/10.1145/1866029.1866072

Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. 2010. Digital image steganography: Survey and analysis of current methods. *Signal Processing* 90, 3 (2010). DOI:http://dx.doi.org/10.1016/j.sigpro.2009.08.010

Christos Danakis, Mostafa Afgani, Gordon Povey, Ian Underwood, and Harald Haas. 2012. Using a CMOS camera sensor for visible light communication. In *IEEE Workshop on Optical Wireless Communications*.

H. Elgala, R. Mesleh, and H. Haas. 2009. Predistortion in Optical Wireless Transmission Using OFDM. In *Proc. of International Conference on Hybrid Intelligent Systems*.

A. Grundhofer, M. Seeger, F. Hantsch, and O. Bimber. 2007. Dynamic Adaptation of Projected Imperceptible Codes. In *Proc. International Symposium on Mixed and Augmented Reality*.

ISO. 2006. QR code 2005 bar code symbology specification. ISO/IEC 18004:2006,. (2006).

K. Kagawa, J. Ohta, and J. Tanida. 2009. Dynamic Reconfiguration of Differential Pixel Output for CMOS Imager Dedicated to WDM-SDM Indoor Optical Wireless LAN. *IEEE Photonics Technology Letters,* 21, 18 (2009).

K. Kamijo, N. Kamijo, and Zhang Gang. 2008. Invisible barcode with optimized error correction. In *IEEE International Conference on Image Processing*. DOI:http://dx.doi.org/10.1109/ICIP.2008.4712185

N. Matsushita, D. Hihara, T. Ushiro, S. Yoshimura, J. Rekimoto, and Y. Yamamoto. 2003. ID CAM: A smart camera for scene capturing and ID recognition. In *Proc. International Symposium on Mixed and Augmented Reality*.

S.K. Nayar and T. Mitsunaga. 2000. High dynamic range imaging: spatially varying pixel exposures. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Richard van Nee and Ramjee Prasad. 2000. *OFDM for Wireless Multimedia Communications* (1st ed.).

Nokia. 2013. LightSpeak - Message in a Photon. (2013). https://research.nokia.com/lightspeak.

Samuel David Perli, Nabeel Ahmed, and Dina Katabi. 2010. PixNet: Interference-free Wireless Links Using LCD-camera Pairs. In *Proc. ACM Mobicom*.

M.S.Z. Sarkera, I. Takai, M. Andoh, K. Yasutomi, S. Itoh, and Shoji Kawahito. 2009. A CMOS imager and 2-D light pulse receiver array for spatial optical communication. In *IEEE Asian Solid-State Circuits Conference*.

Toshiba. 2013. High Dynamic Range. (2013). http://toshiba.semicon-storage.com/ap-en/product/sensor/cmos-sensor/tech-hdr.html.

Jan-Jaap van De Beek, Magnus Sandell, and Per Ola Bö rjesson. 1997. ML estimation of time and frequency offset in OFDM systems. *IEEE Transactions on Signal Processing* 45, 7 (1997).

J. Vucic, C. Kottke, S. Nerreter, K.-D. Langer, and J.W. Walewski. 2010. 513 Mbit/s Visible Light Communications Link Based on DMT-Modulation of a White LED. *Journal of Lightwave Technology* 28, 24 (2010).

Grace Woo, Andy Lippman, and Ramesh Raskar. 2012. VRCodes: Unobtrusive and Active Visual Codes for Interaction by Exploiting Rolling Shutter. In *Proc. International Symposium on Mixed and Augmented Reality*. http://dx.doi.org/10.1109/ISMAR.2012.6402539

S. Yoshimura, T. Sugiyama, K. Yonemoto, and K. Ueda. 2001. A 48 kframe/s CMOS image sensor for real-time 3-D sensing and motion detection. In *IEEE International Solid-State Circuits Conference*.

Wenjia Yuan, Kristin Dana, Ashwin Ashok, Marco Gruteser, and Narayan Mandayam. 2013. Spatially Varying Radiometric Calibration for Camera-Display Messaging. In *IEEE Conference on Signal and Information Processing*.