BitLight: Turning DLP Projections into an Interactive Surface through Bit-level Light Encoding

SONG LIU, University of Minnesota Twin Cities, USA
TIAN HE, University of Minnesota Twin Cities, USA

This paper presents BitLight, a novel paradigm that uses the rapid flashing of a digital light processing (DLP) projector to encode an imperceptible mask temporally that, when sensed by a photodiode, uniquely specifies where the photodiode is located on the projected image. BitLight is inspired by the psychophysical phenomenon that the human visual system (HVS) cannot resolve rapid temporal changes in optical signals, so redundant optical signals could be inserted for tracking with some or little compromise on the original human perceived visual content. BitLight is the first to devise a bit-level temporal encoding to display RGB colors while also embedding tracking signals in a digital fashion. Compared to traditional visible-light-communication (VLC) systems that use frame-level encoding techniques such as luminance changes [10, 25, 26, 32] and alpha channel [16], the bit-level encoding of BitLight makes better use of the ultra-fine temporal division capabilities of DLP projectors to embed a much higher tracking data throughput and thus achieve faster localization speed. With our current prototype hardwares of a low-end microcontroller, cheap photodiodes, and a commercial off-the-shelfDLP projector, evaluation results have demonstrated an average of only 9.5ms to localize the sensor, versus 200ms by a comparison testbed that uses simple frame-level encodings.

CCS Concepts: • **Human-centered computing → Touch screens**.

Additional Key Words and Phrases: Human-computer interface (HCI); Visible Light Positioning; Localization; Visible Light Communication (VLC); Wearable devices

ACM Reference Format:
Song Liu and Tian He. 2020. BitLight: Turning DLP Projections into an Interactive Surface through Bit-level Light Encoding. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 141 (December 2020), 23 pages. https://doi.org/10.1145/3432228

1 INTRODUCTION

Recent years have witnessed an upsurge of interest in integrating visual displays with "touching" capability into electronic devices. Compared with traditional keyboard and mouse systems, its success is largely attributed to its immersive user experience whereby users view and trigger actions on the digital content by simply dragging a finger or stylus.

Existing touching display systems are either expensive or intrude user privacy. Touchscreen panels in Figure 1 are expensive though reliable and have found the widest adoption on electronic gadgets. The other tremendously popular technological category is projector virtual touchscreens, as represented by 2017 Sony Xperia Touch projector in Figure 2. Projector touch-screens are not limited by a fixed-sized physical panel hence have

Authors' addresses: Song Liu, liux3604@umn.edu, University of Minnesota Twin Cities, 4-204B Keller Hall, 200 Union Street SE Minneapolis, MN 55455, Minneapolis, Minnesota, USA, 55455; Tian He, tianhe@umn.edu, University of Minnesota Twin Cities, 4-205 EE/CSci Building, 200 Union Street SE Minneapolis, MN 55455, Minneapolis, Minnesota, USA, 55455.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Association for Computing Machinery.
2474-9567/2020/12-ART141 $15.00
https://doi.org/10.1145/3432228

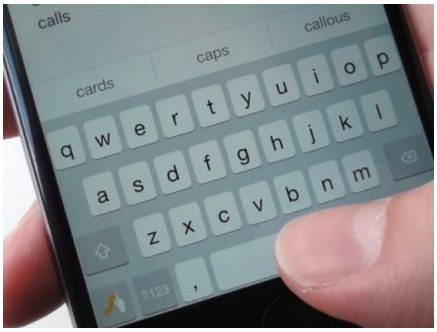

Fig. 1. Capacitive touchscreen

Fig. 2. Sony Xperia Touch[4]

the convenience of projecting to any surface. However the existing technologies use additional near-infrared projectors and cameras for motion sensing, which are both prohibitively costly ($1699 Xperia Touch) and intrude user privacy.

Digital light processing (DLP) projectors, with its unique time-division mechanism in producing colors, appeal to the authors as a perfect candidate for a new paradigm of sensor-display localization that can encode an imperceptible mask temporally to track a light-sensor. Motivated by the information-carrying property of light signals, this paper presents BitLight, shedding light on the possibilities of making use of DLP projectors to embed signals directly and transparently in the light medium for localization. It also gets rid of the need for extra infrared projectors and cameras. In fact, no additional hardware but only an inexpensive light sensor is needed as the receiver-end location tag. The user interacts with the visual content with light-sensor-enabled wearable devices (e.g. stylus or gloves) while viewing the unaffected visual content.

In achieving this, BitLight further proposes to use bit-level binary pulse-width modulation to display distinct colors while embedding signals through bit-level temporal encoding. This is based on the psychophysics phenomenon that the human visual system (HVS) cannot resolve rapid temporal changes in optical signals. Redundant temporal visual stimuli that human eyes cannot discern are thus coded as optical signals for localization, while affecting the perceived visual content very minimally. Advanced techniques including "randomization" and "local-merging" are also proposed to further improve the localization speed.

Contributions of BitLight. The intellectual merits of BitLight are listed as below:

- **New system paradigm: concurrent viewing and sensor-display interaction** BitLight is the first endeavor to turn DLP projector displays into an interactive display through "transparent' information embedding. A light-sensor wearable device (stylus or gloves) can be localized anywhere on the display while the perceived visual content is largely unaffected. The definition of "transparency" is achieving an additional functionality of localization with little or no degradation on the original purpose of displaying visual content.
- **New Technique: bit-level color encoding.** Among all VLC or visible light positioning (VLP) systems, BitLight is the first to propose and prototype the unified approach of bit-level light color encoding. It serves dual purposes: 1) It introduces a systematic and scalable approach to display distinct colors which also transfers much higher data throughput for fast sensor localization. 2) Binary Pulse Width Modulation (PWM) ensures that embedded tracking signals can be accurately read and decoded by the light sensor in a digital manner, difficult otherwise. Our bit-level encoding approach utilizes much higher number of bits, depending on the underlying content, to transfer higher average tracking data throughput per pixel location, as discussed in Section 3.

- **Real Implementation.** Complete prototype with a commercial-off-the-shelf DLP projector proves the feasibility and evaluates the performances of BitLight.

2 BACKGROUND

2.1 Temporal Summation Property of the Human Vision System

The HVS is constantly sampling and integrating the visual stimuli collected during every finite amount of exposure time periodically to form a smooth and stable visual perception of the surrounding [11]. Rapid time-variant optical signals are not discernible beyond a certain frequency, denominated as the critical flicker frequency [2, 29]. In fact, HVS can process and individually perceive up to 10 to 12 images per second, and higher rates are perceived as motion [27]. Rapidly fired optical information are summed up by human nerves in a similar way that the film or digital cameras sensor collect incoming light during its shutter exposure time. Figure 3 and Figure 4 are examples of the digital photos taken at different shutter speed.

Fig. 3. Shutter exposure time: 1/800s [20]

Fig. 4. Shutter exposure time: 1/20s [20]

2.2 The Additive RGB24 Color Model

Humans perceive distinct colors as differently proportioned mixture of three additive primary colors: red, green and blue. RGB24 color model uses 24 bits for the purpose of represent, display and reproduce the theoretically infinitely many distinct colors. It uses three integer triplet (r, g, b) to quantify the shades of three primary color components, representing each primary color component r/g/b with 8-bits in the range $[0, 2^8 - 1 = 255]$. As shown in Figure 5, the RGB24 model can thus represent up to $256^3 = 2^{24} = 16777216$ distinct colors.

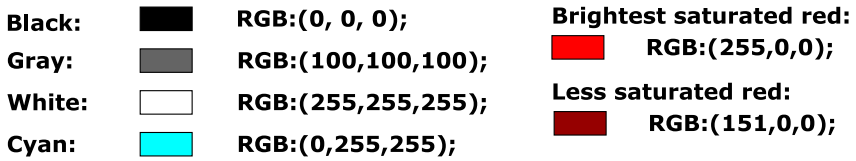

Fig. 5. RGB24 color model uses 256 shades per primary color component

3 BITLIGHT SYSTEM OVERVIEW

3.1 Sensor-display Localization: New Human-computer Interface

Figure 6 shows the overall system conceptual setup. BitLight is a new paradigm that reuses the light medium to achieve the dual purpose of displaying content and location tracking simultaneously, without costly hardwares

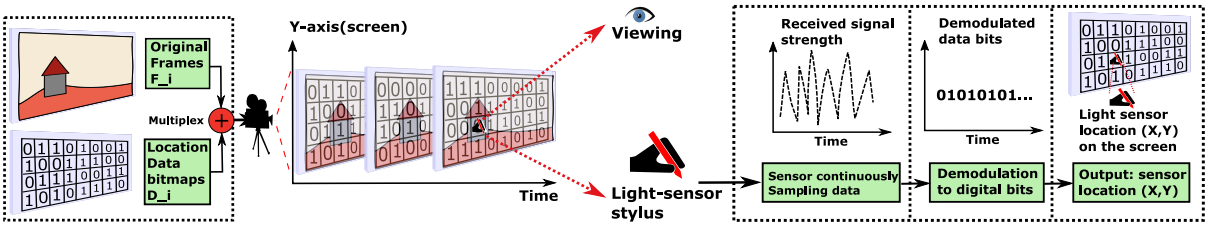

Fig. 6. BitLight: System architecture and operations

like electrically capacitive panels or near-infrared cameras. During the operation, users will hold the photodiode onto the plane of focus of the DLP projector (e.g. projector canvas or wall), since the displayed content is out-of-focus elsewhere due to the light refracting properties of the convex lens in projectors. The photodiode hardware will also need a relatively large receiving angle to at least pick up the light signal. A typical conceivable use case of the system is where a speaker during a presentation can trigger actions on its presentation slides directly by a stylus rod. More details on the prototype hardware implementation can be found in Section 6.

3.2 System Architecture and Operations

Two simultaneous yet distinctive channels are established: one for human viewing and the other for localizing the light sensor as its relative position on the display. The two channels share the same information medium - the displayed content. Image frames F_i with the resolution of $N * N$ are rapidly displayed at the frame rate of f by a DLP projector. A 2D tracking bitmap D_i used for localization is modulated onto each image frame in a human-indiscernible fashion. The light sensor continuously samples the received light signals at its location, demodulates the signals, compares with the 2D tracking bitmaps, and localizes the light sensor S_i to its pixel position (X_i, Y_i) on the screen.

4 BITLIGHT BASIC DESIGN

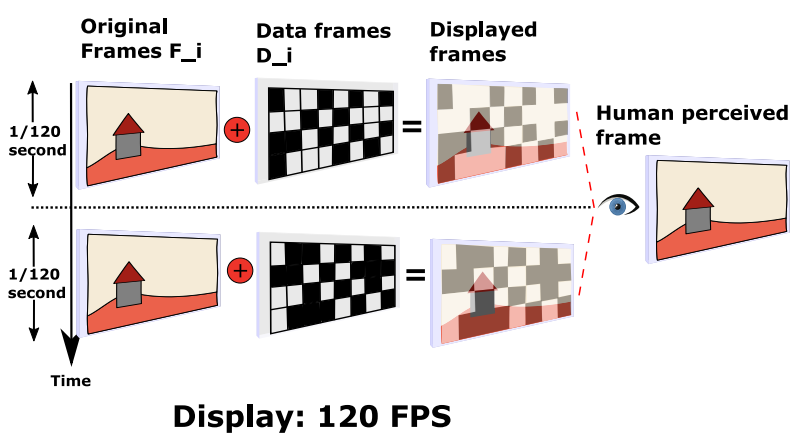

Fig. 7. Frame-level data modulation technique: often used in screen-camera communication systems such as Inframe++ by Anran et al. in [32]

4.1 Limited Throughput of Frame-level Encoding

BitLight is the first to utilize the ultra-fine-granularity time-division capability of DLP projectors for this purpose. Hence understandably, existing VLC encoding schemes no longer work effectively for the new application setup. Existing encodings are mostly frame-level and designed for the different VLC application of screen-camera data communication. They work by multiplexing a 2D data bitmap signal onto each image frame, i.e., it embeds only a fixed number of very few "meta" bit per pixel through means like luminance changes [10, 25, 26, 32] or alpha channel [16] that are indifferent to the underlying displayed content. They are perfectly suitable for camera-screen data communication, whereby the encoded bits on a million+ pixels are captured by a single camera and aggregated as the total communication data throughput. But they are not suitable for screen-sensor localization because the photodiode only collects information at a single pixel location. Reusing the frame-level encoding for our paradigm would lead to extremely slow localization speed.

As an example, to embed tracking data in a human indiscernible fashion, Inframe by Anran et al. [32] is a representative approach that repeats each image twice and modulates two complementary bitmaps to two consecutive frames in Figure 7. Binary bitmaps are multiplexed onto original images by modulating "1" as "increased light intensity" and "0" as "reduced light intensity". The coding scheme of Inframe is implemented on our comparison testbed in Section 6, and evaluation results in Section 7.7 shows that it is 21.05 times slower than BitLight.

Our bit-level encoding approach could utilize a much higher number of bits, which however depends on the underlying content, to transfer a significantly higher data throughput per pixel location. More details are discussed in Section 4.4. Prototype evaluation results in Section 7 shows an fast average localization speed of only 9.5ms, which is the key to open up the door of VLC to the area of sensor-display interaction through rapid localization.

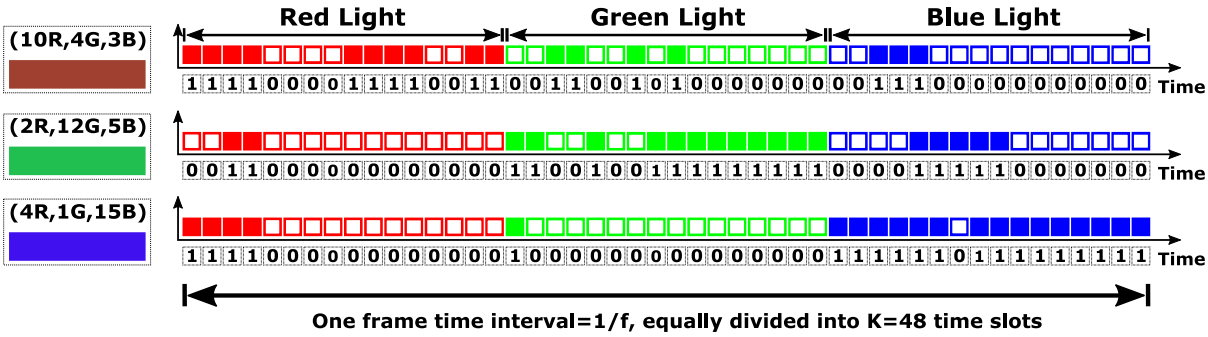

Fig. 8. A new hardware configuration to display colors, the time interval for each frame 1/f is divided into K=48 slots

4.2 Bit-level Binary Pulse-width Color Encoding to Generate Colors

The ultra-fine-granularity time-division capabilities of DLP projectors can be utilized to generate distinct colors scalably, while enabling bit-level temporal encoding with high tracking-data throughput per pixel location. The unique working mechanism of DLP projectors is shown in Figure 8. It generates distinct colors by switching three primary color (red, green and blue) light sources on/off for different portion of time within the time period of a frame. Specifically, the time interval of each frame $1/f$ is partitioned into K equal-length time slots. At any given time, only one of the three RGB principal color light sources is in use, and each principal color light source is in use one third of the time. A bit of "1" switches "ON" the corresponding light source for each $1/f/K$ time

slot, and "0" switches it "OFF". Due to the temporal summation property of the HVS (read Section 2.1), distinct colors are thus created by the frame bit sequences that controls the number of time slots each primary color light source is "ON" during each frame.

Figure 8 illustrates the bit sequences for three different colors with the projector speed of K=48. Notice that the color (10R, 4G, 3B) can be generated by the binary bit sequence 1111000011 110011 (R) 001100101 0000000 (G) 0011100000000000 (B). Since there are more red bits than there are green or blue, the color is more reddish. Each primary color component has $K/3 = 16$ different intensities/shades that together make up different perceived colors. A total of $(K/3 + 1)^3$ distinct colors can be generated, where K is limited only by the hardware device speed. In order to generate as many colors as the original RGB24 color space, the display hardware need to set $K = 768$. In our prototype implementation, our projector hardware is only fast enough to support $K = 24$ at 60fps and supports a total of 729 distinct colors.*The rest of the paper assumes $K = 24$ to ease explanation.* Understandably, the color space of only 729 colors is a major drawback of the prototype. In its current form, our prototype is still far from being a commercially matured product. Hardware enhancements or new configurations from DLP projector manufacturers are needed before it can be commercially successful.

4.2.1 Gamma Correction. Digital camera is a linear device: it receives twice the signal when twice the number of photons are captured. Human eyes are nonlinear in the sense that they perceive twice the light as only a fraction brighter, and increasingly so at higher brightness. Leveraging the higher sensitivity on darker tones of human eyes, the file compression technique of "gamma-encoding" is often applied when recording digital images by redistributing tonal levels closer to human eyes perception [3, 22]. The color encoding method of BitLight is inherently a linear system (gamma = 1) that projects twice the light for twice the number of "ON" time slots. In the future commercialization stage of BitLight, the inverse gamma stage of "display gamma" should be applied during displaying to ensure that the displayed colors matches how it was originally recorded.

4.2.2 Speed Profiles of Modern DLP Projectors. In order to inform readers about the speed profiles and price ranges of the modern DLP projectors, a survey is conducted on the maximum speed limit of DLP devices available on the market. At the core of the trademarked DLP projection technology from Texas Instruments is the digital micro-mirror device (DMD) chipset, which consists of millions of microscopic mirrors which toggles on/off to generate colors in a digital fashion. DMD chipsets decides the theoretical upper limits of the projection speed and resolution, though the final system performance is also affected by other hardware components including 1)light controller chipset, 2)PMIC/LED driver, 3)optical engine, and 4)driver support etc. The maximum binary pattern rate represents the theoretical speed limit of the DMD chipset, though the practically achievable speed also depends on other factors including flash memory size, data bandwidth, driver support and input source format etc. Our survey results are listed in Table. 1. As an example, our prototype DLP projector model DLPDLCR4500EVM could only be programmed to generate K=24 at 60fps despite of the chip's theoretical speed limit of 4225Hz binary pattern rate. If everything being same, the high-end chipset DLP7000 from the same manufacturer is $32552/4225 = 7.7$ times faster than our current prototype chipset DLP4500. This gives a theoretically equivalent value of $K = 24 * 7.7 = 184$, which is capable of generating a colorspace size of $(K/3+1)^3 = 242192$ distinct colors. It should be noted that this is a only a theoretical estimate since the authors do not own one of the DLP7000 projector and many details are also missing from its technical sheets. Still, it should provide some confidence in faster hardwares to help scale up the K value and the colorspace of BitLight design in the future.

4.3 Signal Encoding, Decoding and Localization

This section explains the encoding and decoding of tracking signals, and the subsequent stage of sensor localization by mapping the decoded bit sequence with the encoded bitmaps.

Table 1. Survey on the maximum projection speed and price of modern DLP devices.

DLP DMD chipset	Product	Max binary pattern rate	MSRP
DLP3010LC	DLP3010EVM-LC	2500 Hz	$799
DLP3310	DLPDLCR3310EVM	N/A	$849
DLP4710	E4750RGBLC	1440 Hz	$1199
DLP4710LC	DLP4710EVM-LC	2500 Hz	$1299
DLP2010NIR/DLP4500NIR	DLPNIRNANOEVM	2880/4225 Hz	$999
DLP4500	DLPDLCR4500EVM	4225 Hz	$1299
DLP5500	N/A	5000 Hz	$403
DLP6500FYE	DLPLCR6500EVM	11574 Hz	$1999
DLP9000	DLPLCR9000EVM	9523 Hz	$5749
DLP9000X	N/A	14989 Hz	$4449
DLP9500	DLPLCR95EVM	23148 Hz	$3999
DLP7000	DLPLCR70EVM	32552 Hz	$1799

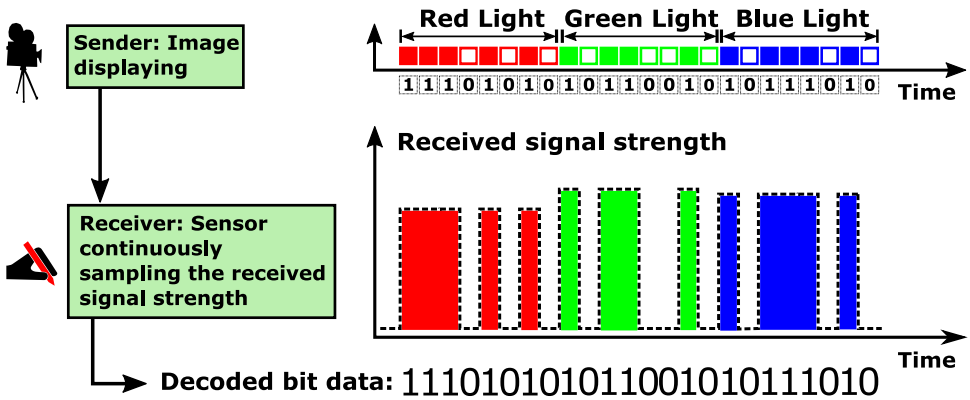

Fig. 9. Digital demodulation of data bits: K=24

4.3.1 Encoding of Bit Signals. As described in Section 4.2, the DLP projector reads binary pulse-width modulated sequences, stored in the format of preprocessed video clips, to generate distinct colors on display by controlling the ON/OFF state of the three light sources on the time domain. Tracking bit signals are thus encoded as "ON/OFF" of the light source at any given time. As Figure 9 shows, the tracking bit sequence "11101010 (R) 10110010 (G) 10111010 (B)" is encoded temporally as: ON ON ON OFF ON OFF ON OFF ON OFF ON ON OFF OFF ON OFF ON OFF ON ON ON OFF ON OFF at the pixel location.

4.3.2 Decoding of Bit Signals. The receiver light sensor, coupled with a micro-controller, continuously samples its received light signal strength, and digitally decode the binary state of existence/absence of light as 1/0 bit, as shown in Figure 9. The common conclusion from the state-of-the-art VLP systems [18, 30, 34] that the indoor office/lab lighting condition is highly stable under normal operation is confirmed and adopted. The room lighting can be pre-measured and subtracted from the analog sensor readings. The power line blinking frequencies of 100/120Hz (doubled by the rectifier circuitry) can also be easily compensated due to its high regularity. Our testbed

experiments have confirmed that its influence is trivial, since its amplitude change is small when compared with the much higher projector brightness.

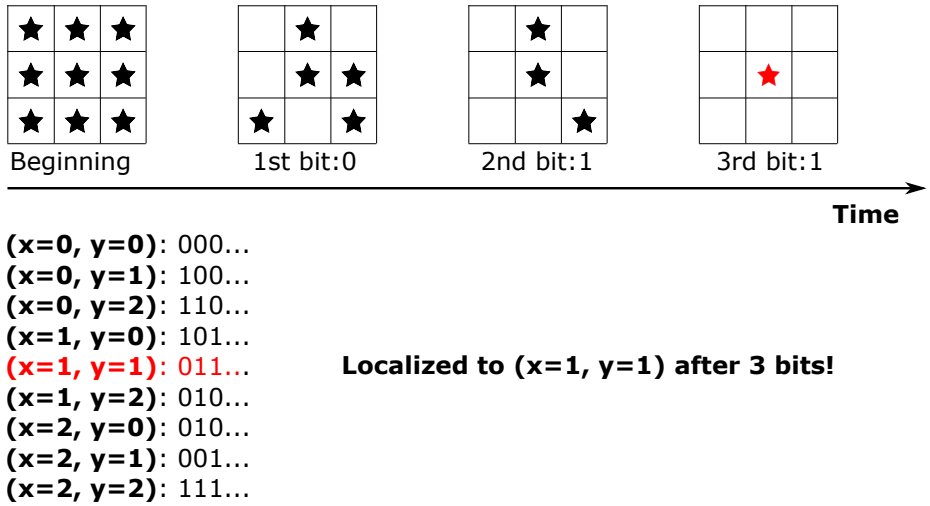

Fig. 10. Example with a resolution of 3*3 to demonstrate the localization process.

4.3.3 Calculating the Sensor Location. The decoded bit sequence is compared with the modulated bitmaps to calculate the sensor location on the display. During system calibration, the receiver can be easily synchronized with the projector by playing a pre-defined bit pattern. The localization process is an iterative converging process, whereby the pool of candidate sensor locations shrink bit-by-bit until its location can be unique identified. Figure 10 is a miniature example with only $3 * 3$ resolution to illustrate this process. The encoded bitmap data at each pixel is known beforehand as follows: "(x=0, y=0): 000...; (x=0, y=1): 100...; (x=0, y=2): 110...; (x=1, y=0): 101...; (x=1, y=1): 011...; (x=1, y=2): 010...; (x=2, y=0): 010...; (x=2, y=1): 001...; (x=2, y=2): 111...". The receiver compares its decoded bit sequence with the bitmap until the location of the sensor could be uniquely determined. At the beginning, the sensor could be located at any of the 9 pixel locations. After the first bit of "0" is decoded, the searching range reduces to a smaller set of possible locations: "(x=0, y=0), (x=1, y=1), (x=1, y=2), (x=2, y=0), (x=2, y=1)". After the second decoded bit of "1", the searching range is further reduced to "(x=1, y=1), (x=1, y=2), (x=2, y=0)". After the third decoded bit of "1", the sensor location is finalized at "(x=1, y=1)". It takes a total of "3 bits" time to successfully localize the sensor.

4.4 Randomization Technique: Visually Equivalent Binary Sequences

To achieve the goal of dual-channel viewing and sensor localization, the design should minimize the compromise on the human perceived visual content as much as possible. At the same time, the transparent design philosophy of BitLight requires that no restriction be imposed on the underlying visual content displayed. This poses yet another design challenge: if a large number of pixels share an identical color for a prolonged number of frames, the localization speed is slower at those regions of the display compared to others. This spatially in-homogenous distribution of localization speed is illustrated in Figure 27 and Figure 28, whereby some regions of the apple orchard video consistently carries the same content.

According to the temporal summation property of the HVS (Section 2.1), multiple 1/0 sequences can control the three primary color lights differently yet reproducing an equivalent human perceived color. As an example,

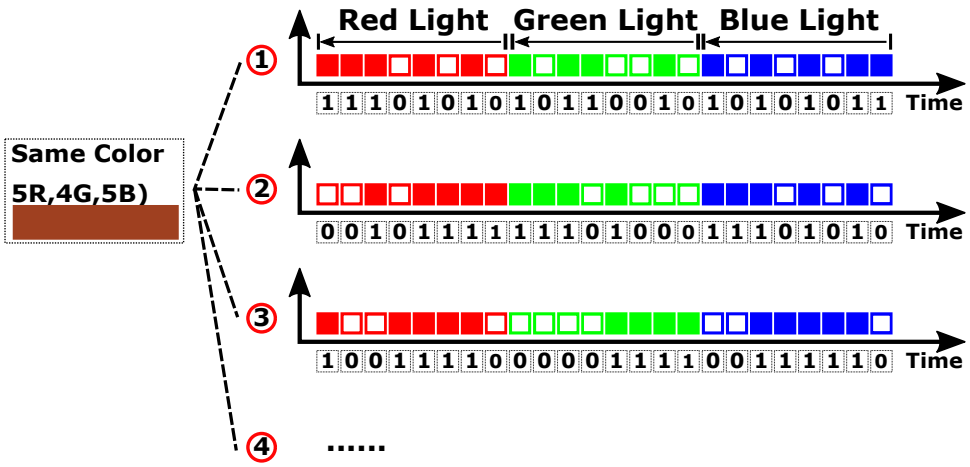

Fig. 11. Color $(5R, 4G, 5B)$ can be generated with different 1/0 sequences(K=24)

Figure 11 demonstrates that the same human perceived color $(5R, 4G, 5B)$ can be reproduced by ${}^{8}C_5 * {}^{8}C_4 * {}^{8}C_5$ [1] different binary sequences, including but not limited to the following sequences: 1) 11101010 (R) 10110010 (G) 10101011 (B), 2) 00101111 (R) 11101000 (G) 11101010 (B), and 3) 10011110 (R) 00001111 (G) 00111110 (B). Thus the system could modulate large data throughput by *randomly* picking any visually "equivalent" binary sequences to generate a same color. The receiver captures the exact bit sequences and recognizes them as different tracking data. We coin this as "randomization technique". The randomization technique ensures fast localization speed and alleviates the problem of spatial inhomogeneity of color distribution as evaluated in Section 7.8.

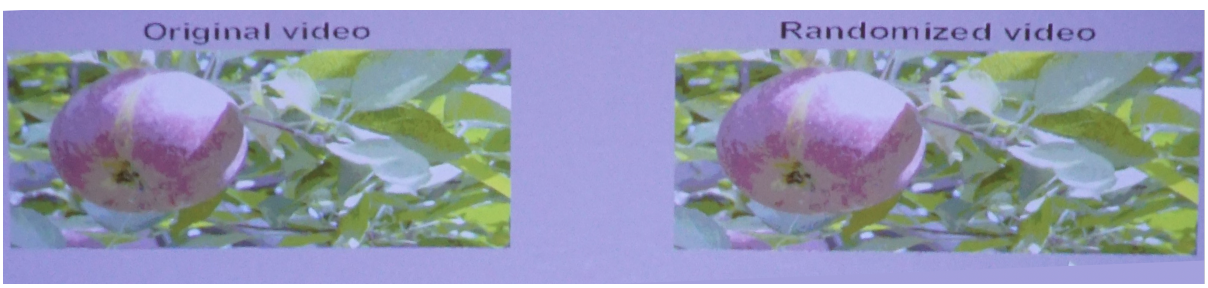

Fig. 12. Side-by-side comparison of the original unprocessed videos and processed videos with randomization techniques captured by DSLR camera.

In order to give the reader a visual tour, a side-by-side comparison of the original unprocessed videos and processed videos with randomization techniques is recorded in Figure 12, captured with a Sony A6000 Mirrorless DSLR Camera with 16-50mm lens. The images are recorded in a manual-mode with exposure time of 1/13 seconds, close to the equivalent exposure time of the human vision system [27] (read Section 2.1).

[1] ${}^{N}C_k$ is the combination notation for N choose k.

4.5 Information Carrying Capability of Different Colors

It could be inferred from the randomization technique that different colors have a varying capacity of carrying tracking data. Each RGB color (r, g, b) can be reproduced by a different number of visually equivalent binary sequences, quantified by ${}^8C_r * {}^8C_g * {}^8C_b$. Apparently, the color $(4R, 4G, 4B)$ can carry the most tracking data per pixel per frame, much larger than $(2R, 0G, 0B)$, since ${}^8C_4 * {}^8C_4 * {}^8C_4 >> {}^8C_2 * {}^8C_0 * {}^8C_0$.

4.5.1 All Bright and Black Colors. There is however an extreme case whereby the entire display area is the "brightest white" color 8R8G8B or the "darkest black" color 0R0G0B for prolonged time, whereby their number of equivalent binary sequences is ${}^8C_8 * {}^8C_8 * {}^8C_8 = {}^8C_0 * {}^8C_0 * {}^8C_0 = 1$. To handle this extreme case, the system could simply rule out these two colors. We could compromise slightly on the color space by replacing the fully white color 8R8G8B with 7R8G8B, and fully black color 0G0G0B with 1R0G0B. Admittedly, this will reduce the colorspace size and also the contrast ratio, defined as the ratio of the luminance of the brightest white color to the darkest black color. Similar to OLED displays, the DLP displays of BitLight originally have a theoretically "infinite" contrast ratio since a perfect black is generated by emitting no light. Now since the darkest and brightest white are no longe available, the contrast ratio is reduced to $K/3 : 1$. The reduction in colorspace can be ignored for fast hardwares that can afford large values of K and is capable of generating millions or billions of distinct colors. The loss in contrast ratio is practically fine, since the human perception of black is also relative, affected by many factors including the room lighting condition.

5 LOCAL BIT-EXCHANGE: FURTHER SPEED UP LOCALIZATION

This section introduces a spatial filtering technique to further increase the localization speed, particularly for projectors with ultra-high resolution densities.

5.1 Intuition

This improvement technique is based on the psychophysical phenomenon that human visual perception has a limited spatial resolution in discriminating fine details. Studies have shown that the typical spatial resolution of human eyes is $0.07°$ or approximately 1.2mm from 1 meter watching distance [5]. With rapid hardware advancement, modern display devices could easily support tens of mega-pixel resolutions, far exceeding the human retina discernibility. This gives the opportunity to exchange color content between nearby pixels by trading off the perceived resolution slightly. Understandably it will introduce extra visual artifacts, however, the visual artifact is increasingly minimized with higher resolutions and from a longer viewing distance. The visual artifact at different viewing distances and resolutions is evaluated in Section 7.3 and Section 7.4, the results of which have verified this claim. The objective is thus to manipulate the number of appearance of each color in each frame, such that the color distributions are ideally proportionate to their data-carrying capacity as much as possible.

5.2 The Local Bit-exchange Spatial Filter

We define a bit-exchange spatial filter, which groups every M nearby pixels, whereby the group's sum of the count of "ON" bits of each r/g/b color within a frame can be freely exchanged. Figure 13 shows an example of $M = 4$ whereby every 4 nearby pixels form into one quadruplet. The quadruplet on the top left corner of the image originally have 2R, 1R, 1R and 4R respectively, summing together to 2+1+1+4 = 8 time slots of "ON" state of red light. Hence after applying spatial filter, the quadruplet can be replaced by one of many combinations: 1)2R, 2R, 2R, 2R 2)3R, 1R, 2R, 2R 3)3R, 1R, 3R, 1R ... which all sum up to 8R. The combination used to substitute the original values are selected randomly according to a probabilistic distribution, which is calculated using quadratic programming as explained in the next subsection. Compared with the randomization technique in Section 4.4, the spatial filter technique further considers every M local pixels instead of treating every pixel individually.

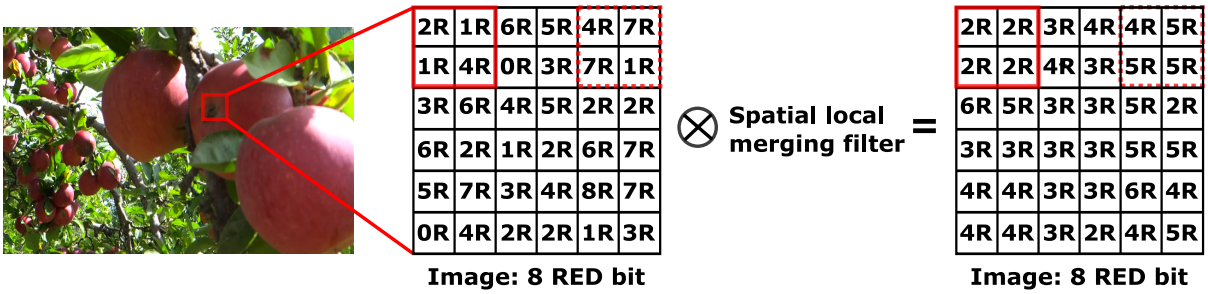

Fig. 13. Applying the spatial filter of local bit-exchange to the RED bits of a frame

Note that the spatial filter window is moved horizontally and vertically only with an integral step size of the square root of M. Hence, we cannot move the spatial filter in the figure to the right by just one pixel at a time. The bit-exchange spatial filter should be used in a mathematically smart and calculated manner in processing the images, so as to optimize the distribution of colors in each frame. Figure 14 illustrates the effects of local bit-exchange with varying values of M. Understandably, the image quality degradation increases with M.

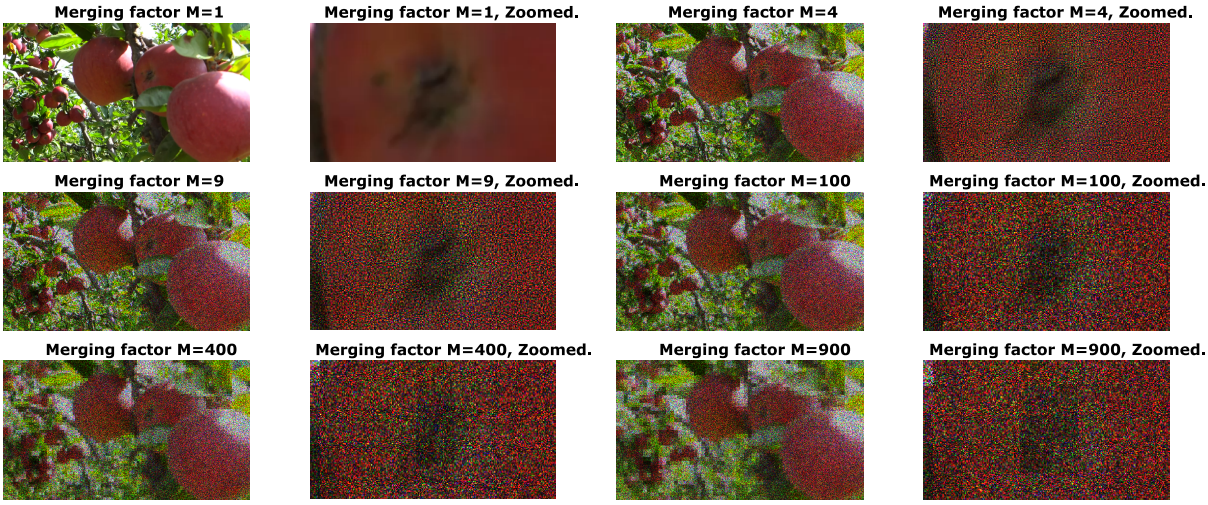

Fig. 14. Local bit-exchange spatial filter with a different merging factor M.

5.3 Quantitative Analysis of the Localization Speed

The average localization speed could be quantized with the assumption that the sensor location is unknown and has an even likelihood to be potentially located at any of the N_{total} pixels on the display. As described in the evaluation Section 7.7, the localization speed of our design is often faster than a frame. Hence we examine at the finer granularity of every $K/3 = 8$ bits instead of every frame to gain better insights.

After the first $K/3 = 8$ bits of the same primary color have been played, the number of candidate searching space will be shrunk from N_{total} to an average of $\sum_{i=0}^{8} \frac{N_i^2}{{}^8C_i} \cdot \frac{1}{N_{total}}$ as in Equation. 1 (see Section 4.4). Increasing

the average localization speed is thus effectively equivalent to reducing the objective function in Equation. 1, in other words, maximizing the converging rate of localization.

$$\boldsymbol{Min} \sum_{i=0}^{8} \frac{N_i}{{}^8C_i} \times \frac{N_i}{\sum_{j=0}^{8} N_j} = \sum_{i=0}^{8} \frac{N_i^2}{{}^8C_i} \times \frac{1}{N_{total}} \Rightarrow \boldsymbol{Min} \sum_{i=0}^{8} \frac{N_i^2}{{}^8C_i} \tag{1}$$

, where $N_{total} = \sum_{j=0}^{8} N_j$ is the total number of pixels and omitted as a constant. N_i is calculated as the number of pixels on the display that have i ($i \in [0, 8]$) bits "ON" as in Equation. 2. For example, an image pixel with the bit sequence 11011000 (R) has four red bits "ON" and increments N_4 by one.

$$\boldsymbol{N} = [N_0, N_1, N_2, N_3, N_4, N_5, N_6, N_7, N_8]^T \tag{2}$$

The optimization problem formulated above of finding probabilistic distribution of quadruplet splits can be easily solved by classical convex quadratic programming methods. Readers can referred to advanced polynomial-running-time approximate algorithms [36, 37].

6 SYSTEM IMPLEMENTATION

Hardware implementation The prototype uses a DLP projector of model DLPLCR4500EVM from Texas Instruments, which supports up to 120fps (but only 60fps is used in implementation due to Windows OS NVidia driver support), 24-bit (K=24) and a native resolution of 912x1140. The video images displayed by the projector is fed from a desktop computer configured with Nvidia GTX1070, Intel i7-4790 processor and 16GB memory. The receiver side as in Figure 17 are several light sensors (Honeywell SD3410-001), each serially connected to a micro-controller unit (Arduino Uno R3) through a 1K ohm resistor. The micro-controller continuously measures the voltage across the resistor as an indicator of the light intensity it receives. Honeywell SD3410-001 has a relatively small form factor (sensitive area < 1-millimeter square), wide sensitivity range, and a wide acceptance angle of -90° to 90°. In choosing the photodiodes, special emphasis should be given to the small sensitivity region. As shown in the evaluation Section 7.5, small sensitivity region leads to finer localization granularity. Its short rise/fall time of 75μs also gives it a high theoretical sampling rate up to 13.3kHz, which is already an overkill for our prototype that requires only $2 * 60 * 24 = 2880Hz$. According to Nyquist theorem, analog signals should be periodically sampled at a rate at least 2X the highest frequency needed to record. Hence the sensing speed of chosen photodiode hardware should be at least $2 * f * K$, where f is the display frame rate. Both the projector and the receiver sides are connected to the same computer, which displays the processed images and calculates the sensor location based on the decoded light signals. Two prototypes of the receiver end are implemented as shown in in Figure 17 and Figure 19 and Figure 20 according to the circuit schematic in Figure 18. Second prototype has a magnet attached to the white board for the ease of carrying out extensive experiments at different locations.

Software implementation An ad hoc C++ DirectX 11 graphic display engine is developed and calibrated to work with the projector driver such that the 0/1 sequences of the image frames decides to the ON/OFF states of the RGB light sources inside the projector to achieve time-division.

Wearable smart gloves As a demonstration of possible use-case, a receiver end in the form of wearable glove is built as in Figure 16. The glove has two light sensors sewn into the gloves at the thumb and index finger positions. The wearable gloves is not yet a matured wearable product suitable for ubiquitous usage, but it should inspire future researchers/engineers to come up with better hardware setup or designs to overcome current usability limitations. Section 8 also discusses the practical inconveniences of SmartLight as a wearable system compared with computer vision systems.

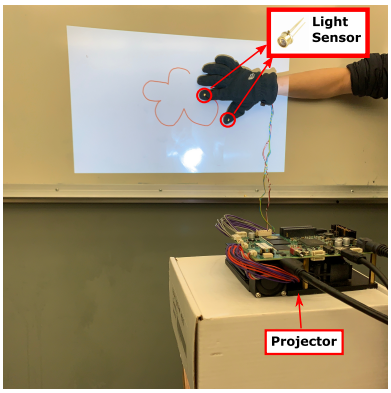

Fig. 15. System setup

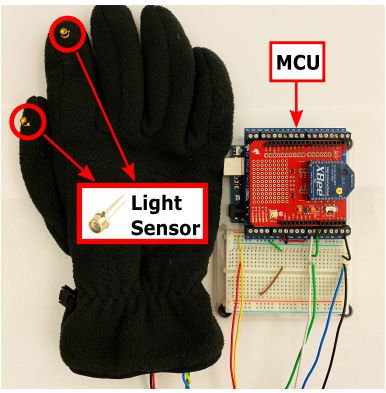

Fig. 16. Wearable gloves

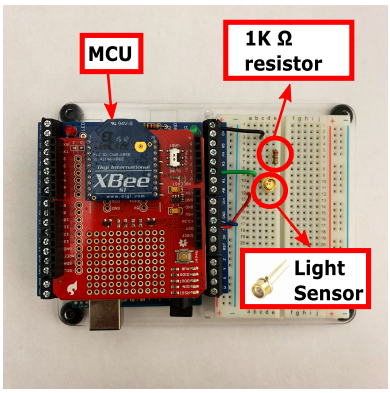

Fig. 17. 1st prototype

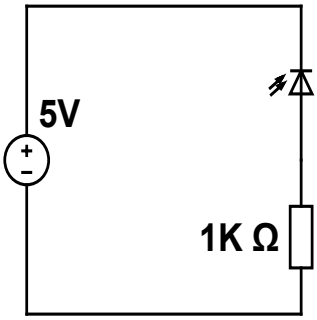

Fig. 18. Circuit

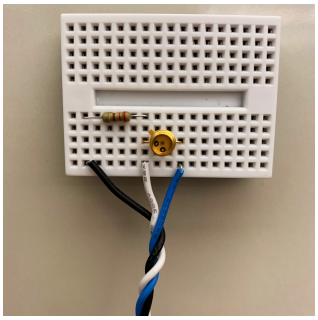

Fig. 19. 2nd prototype

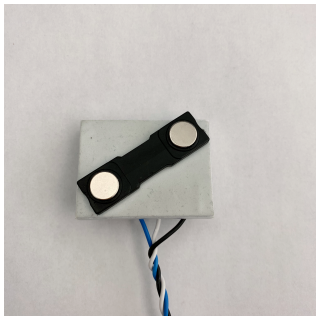

Fig. 20. 2nd prototype

7 SYSTEM EVALUATION

The prototype is evaluated on the perceived visual quality degradation, localization granularity and average/worst case localization speed with respect to different design parameters.

Fig. 21. Twelve video clips used for all evaluation experiments representative of different video characteristics including: high/low contrast and fast/slow motion.

7.1 Video Clips Used for the Evaluations

The "transparent" design philosophy of BitLight aims at being non-selective on the underlying visual content displayed. Kept this in mind, a selection of 12 license-free video clips are gathered and used as the underlying video sources, as shown in Figure 21. They should be representative of different characteristics including: high contrast (sunset, sunrise, lake), low contrast (golf, fish, apple orchards), fast motion (city traffic, driving, rocket launching), slow motion (fish, apple orchards, golf). In order to evaluate on the extreme case that the entire video consists of one single color from the beginning to end, we artificially crafted a pure gray color video with a single color of 4R4G4B or single sequence of "00001111 (R) 000011 11 (G) 00001111 (B)" throughput the video. The 12 video clips are used and evaluated in every one of the following evaluation experiments with the corresponding performance reported.

7.2 Experimental Setup for Subjective Human Perception Assessment

A user study are carried out to verify that the subjective human perceived visual degradation is within acceptable range across different design parameters. The study involves 10 volunteers: 7 males and 3 females, aged from 22 to 31, and half of whom wear glasses. The original video and the revised video processed by the algorithms of BitLight, are played on the same hardware platform with identical settings of $9^3 - 2 = 727$ color space. Participants are asked to compare both videos and report their subjective visual experience. A common metric used in similar visible light communication systems such as Inframe by Anran et al. [32] is adopted: the score range of 0 to 4 is used, which represents "no video degradation", "very little degradation", "some degradation", "evident degradation" and "significant video artifact" respectively.

7.3 Evaluating the Subjective Perception with Respect to Video Sources

Subjective human perception are evaluated on the 12 underlying video clips in Figure 21 while keeping all other parameters unchanged at 912*1140 video resolution, 20cm projection distance and 3m viewing distance. The results in Figure 22 confirms the design goal that different video streams have some influence on the user viewing experience. But the visual artifact is within the acceptable range at a normal viewing setup. Gray video and color strips video have reported more visual degradation partly because the visual difference on a much smaller number of colors are more noticeable to human eyes. The result is also in accordance with our expectation that using the advanced bit-exchange speedup technique in Section 5 will introduce more noticeable visual artifact, and it increases with the bit-exchange factor M. This effect however is alleviated and can be conceivably got rid of when the hardware can afford a high pixel density, or high resolution on a small display area.

7.4 Evaluating Subjective Perception with Respect to Distances/Resolutions

This experiment evaluates the average perceived video quality degradation at different projection distances (larger projection distances induce larger projection areas) and video resolutions at the same viewing distance of 3 meters without bit-exchange techniques. Figure 23 shows that the average video quality degradation decreases with higher resolution and lower display distances. Visual degradation decreases at smaller projection distances, because it leads to a higher pixel density with a smaller display area, and potential artifacts are visually "merged away" spatially. Since modern projection devices often have at least 1080p or even 4K resolution, we conclude that the induced visual perception degradation is safely acceptable to be considered practical.

7.5 Evaluating the Localization Accuracy or Granularity

The localization granularity is theoretically and experimentally verified to be only limited by the smaller value of 1) the sensitivity area of photodiode, and 2)the physical pixel density of the projector display. Our choice of photodiode Honeywell SD3410-001 TO-46 Metal has a very small sensitive area less than 1-millimeter square,

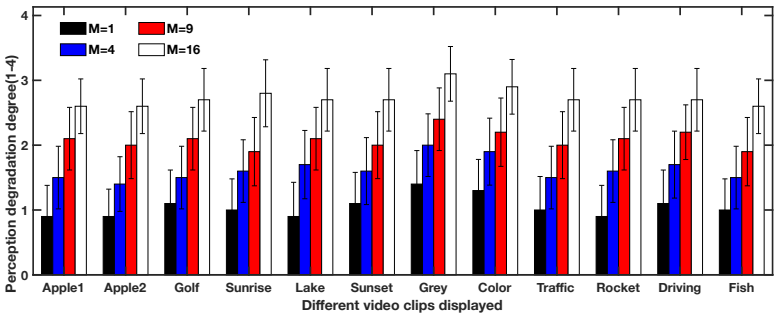

Fig. 22. Human perceived quality degradation V.S. local bit-exchange factor M and video sources. Notice that local merging is not applied when M=1

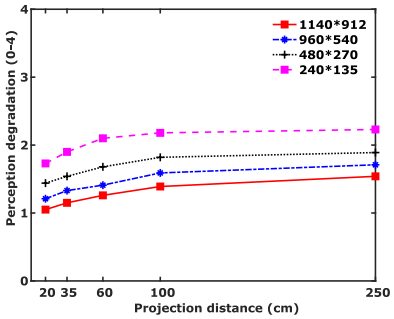

Fig. 23. Human perceived quality degradation V.S. distances and video resolutions

which provides very high granularity. The "Flat Window Can Package" of the photodiode hardware helps to increase signal-noise-ratio but do not collect noticeably more light rays not focused at its location, as verified during the experiments. Note that the photodiode is placed on the plane of focus of the projector, hence it does not need an extra lens to collect the already focused light ray at its location. A comfortable short throw viewing distance of 60cm are adopted for the experiments, and a consistent video resolution of 76*42 are used for all accuracy and speed evaluation experiments. BitLight achieves an average of 2.1mm accuracy which is close to half of 6mm: the physical length of each pixel.

Several state-of-the-art VLC localization systems listed in the following are implemented and compared with BitLight in the same lab setting. Okulis is the only exception, as the authors do not have the resources or expertise to replicate and calibrate such an ad-hoc and sophisticated hardware system. Its reported performance is directly reused.

- **Epsilon (NSDI'14) [14]**: uses a multi-lateration method and includes the impact of incidence angle in its Lambertian received-light-signal-strength (RSS) propagation model.
- **Okulis (Mobicom'15) [39]**: uses a low-power LED that transmits lights and two light sensors to locate user's finger on the 2D plane by examining the light reflections from the finger. It achieves around one-centimeter location precision.
- **SmartLight (Sensys'15) [18]**: uses a single modified LED light to localize light sensors in 3D. The underlying technique is to exploit the light splitting properties of convex lens to create an one-to-one mapping between a location and the set of orthogonal digital light frequency signals receivable at a location.
- **Spotlight (Sensys'05) [30]**: generates a sequence of timed spatio-temporal events of on/off patterns and localizes a light sensor by the timestamps of light shedding on it.

Figure 24 compares the cumulative distribution functions of the localization error of BitLight with the state-of-the-art light-based localization approaches including SmartLight, Okulis, Spotlight and Epsilon. Notice that BitLight is the only system that aims to achieve unaffected viewing and localization simultaneously. Among all methods, BitLight and SpotLight have achieved superior performance as its granularity is theoretically only bounded by the receiver physical form-factor and the video pixel resolution. The experiments also eliminate the concern on the pixel boundary being a blurry region for localization, primarily due to the fact that 1)a photodiode with extremely small sensitivity region can be picked as receiver, and 2)photodiode operates on the focus plane of the projector. Okulis also achieves excellent performance even though its implementation requires significant effort in system calibration.

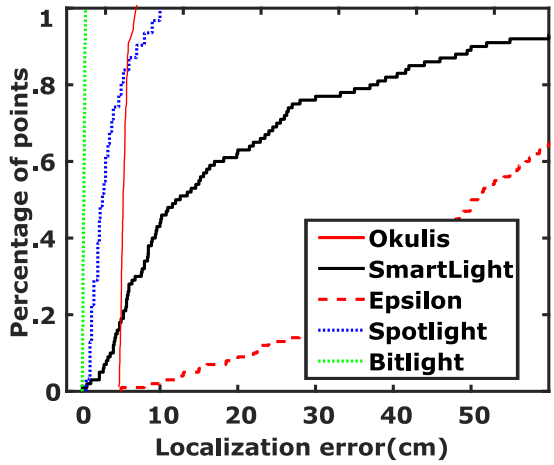

Fig. 24. Evaluating the CDF of localization error V.S. the state of the art

7.6 Experimental Setup for Speed Evaluations

All speed evaluation experiments are carried out on the same testbed with $76 * 42$ video resolutions at 60fps, across all 12 video clips in Figure 21 at 20 random locations. Localization time is capped at 50 frames or 0.83 seconds, beyond which is considered impractically slow. "Frame-level" is the traditional frame level technique which encodes 1 bit per pixel per frame. "BitLight-Original" is displaying original unprocessed images on BitLight as if it were designed for BitLight, and it is included only for reference. "BitLight-Naive" is the simplest version of BitLight without randomization techniques applied. "BitLight+Randomization" is a better version of BitLight with randomization technique applied but without bit-exchange technique applied. "BitLight+Randomization+bit exchange M=4" is the fully fledged version with randomization and bit-exchange applied with a factor of $M = 4$.

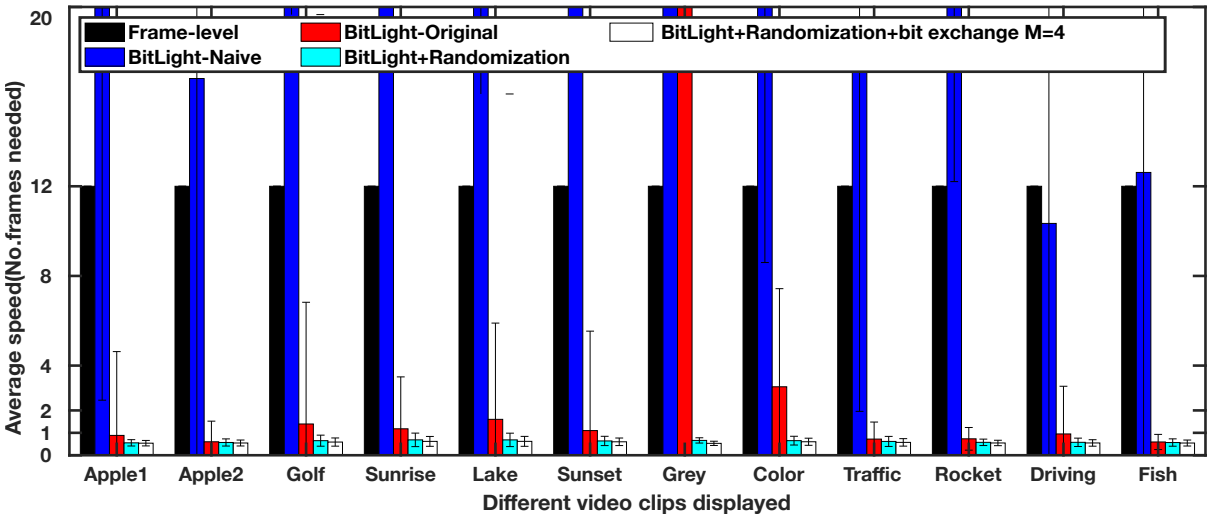

Fig. 25. Average localization speed V.S. video sources, with/out speedup techniques.

7.7 Evaluating the Average Localization Speed with Bit-level Encoding

The frame-level encoding by artificially scaling the perceived light intensity higher or lower while keeping its color proportions as close to the original color, similar to Inframe by Anran et al. [32] and Uber-in-light by Mostafa et al. [10] is implemented as a comparison system. Results show that the fully fledged version "BitLight + Randomization+bit exchange M=4" localizes the sensor with an average of 0.57 effective frames or 9.5ms on the 12 videos, compared with 12 effective frames or 200ms with frame-level encoding. This is 21.05 times faster! "BitLight + Randomization" localizes in 0.62 frames on average, which is still 19.3 times faster. "BitLight-Naive" is even worse than "Frame-level" with an average of 28.5 frames, and this is understandable because a large portion of the videos carries the same color. This over-simplified version of BitLight does not use visually equivalent bit sequences to resolve this problem (read Section 4.4). However it should also be noted that our current hardware, with $K = 24$, could only support $729 - 2 = 727$ color space. The stated speedup is currently at the cost of colorspace, which makes it far from a commercially matured product. A much faster projector is needed to achieve commercially viable color space.

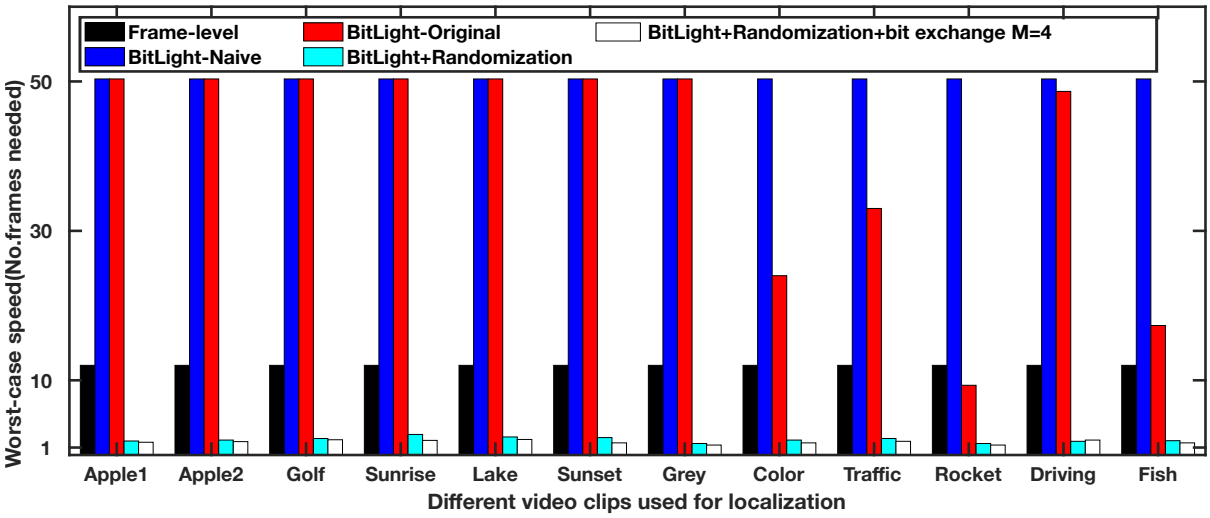

Fig. 26. Worst-case localization speed V.S. video sources, with/out speedup techniques.

7.8 Evaluating the Spatial Distribution and Worst Case Localization Speed

BitLight imposes no restriction on the underlying visual content displayed. Hence if a large number of pixels carries identical colors for a prolonged time, it brings out the problem of in-homogenous distribution of localization speed spatially across the display. Randomization technique described in Section 4.4 will understandably alleviate this problem, since it randomly choose from multiple visually equivalent binary bit sequences to replace the original bit sequence. Figure 27 serves as a good visual demonstration of the speedup after randomization technique. It is a snapshot of one of the evaluated apple orchard video listed in Figure 21. Notice that due to sunlight reflection, a large portion on the apple and leaves carries bright white colors for up to 15 seconds in the original video. Hence it takes much longer time to localize at those regions.

Experiments have reported a significant improvement on the in-homogenous distribution of localization speed spatially on all 12 videos. "BitLight+Randomization" localizes with an averaged mean of 0.62 frames (10.3 ms) and standard deviation of 0.2 frames (3.33 ms), versus "BitLight-Naive" with an averaged mean of 28.5 frames (0.475

s) and standard deviation of 14.7 frames (0.245 s). The worst case localization speed of "BitLight+Randomization" is an average of 2.1 frames (35 ms) versus 50 frames (0.83 seconds) by "BitLight-Naive" on 12 videos.

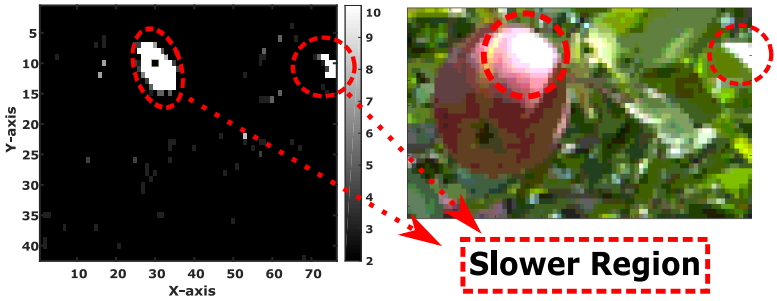

Fig. 27. The spatially in-homogenous distribution of localization speed without randomization for BitLight-Original, measured by the number of frames.

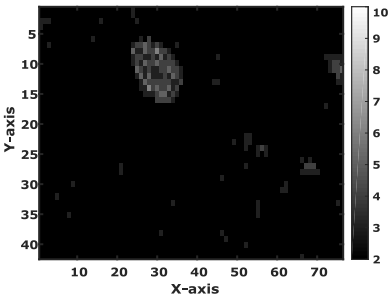

Fig. 28. The spatial distribution of localization speed after randomization

The pure gray color video in Figure 21 has observed the largest speedup and is worthy of special attention. "BitLight-Naive" have reported a mean and max of 50 frames (0.83 seconds) and zero standard deviation. "BitLight+Randomization" is able to reduce that to a mean of 0.66 frames (11 ms), a standard deviation of 0.12 frames (2 ms) and a worst-case max of 1.54 frames (25.7 ms). This matches our expectation that the purely gray color video will otherwise take infinitely long to localize.

7.9 Evaluating the Localization Speed with Bit Exchange Technique

We compare the localization speed with and without the bit exchange image preprocessing techniques described in Section 5. Results in Figure 25 show that "BitLight+Randomization+bit exchange M=4" localizes in an averaged mean of 0.57 frames (9.5 ms), standard deviation of 0.15 frames (2.5 ms) and worst-case max of 1.75 frames (29.2 ms), versus "BitLight + Randomization" with a mean of 0.62 frames (10.3 ms), a standard deviation of 0.20 frames (3.3 ms) and a worst-case max of 2.05 frames (34.17 ms). It is noteworthy that the pure grey video has seen the most speedup of 25% in mean localization speed. Golf, sunrise and lake video have observed an average of around 10% speedup on mean localization speed. This is understandable as the gray video is crowded with only a single color, hence bit-exchange technique takes advantage of a wider spectrum of available colors, drastically boosting up the overall localization capability. Sunrise and sunset have observed the largest decrease in worst-case max localization by a factor of 40% and 43% respectively. This is because high-contrast videos have a lot of very bright and dark colors that can be optimized.

7.10 Evaluating the System Scalability at High Resolutions

The average and worst case system speed are simulated at different resolutions of $42 * 76$, $168 * 304$, $336 * 608$, $504 * 912$ and $1080 * 1920$ across all the 12 videos. The simulations complement our real testbed evaluations and have enabled system evaluations up to the original video resolution of 1080P, and at a scale not feasible or too time consuming otherwise. Each video is evaluated for a total of 1000 times at 50 random pixel locations across 20 random temporal starting points. Results in Figure 29 and Figure 30 show that the localization speed does not increase significantly with the resolution count. In fact, it follows a logarithmic running time complexity with respect to the video resolution. Increasing the resolution by 649 times from $42 * 76$ to $1080 * 1920$ only increases the average localization time by 103.7% from 13.96 bits (0.582 frames or 9.7 ms) to 28.4 bits (1.18 frames or 19.8 ms). The worst case localization time increases by only 112% from 31.4 bits (1.3 frames or 0.022 s) to 66.5 bits (2.77 frames or 0.046 s). This is in accordance with our analysis in Section 4.3.3 that the localization process

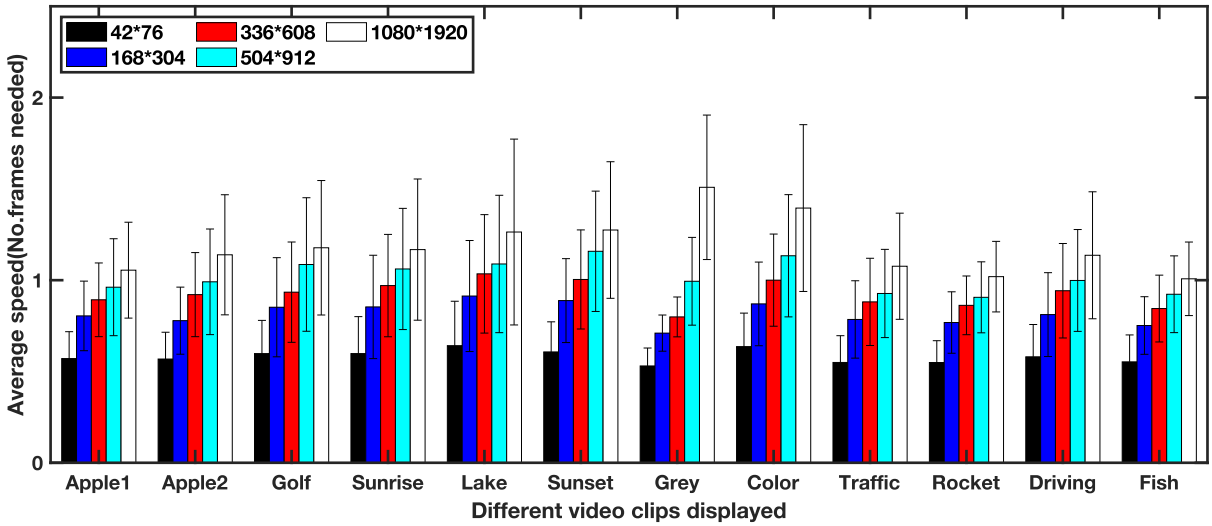

Fig. 29. Evaluating the average localization speed at different video resolutions.

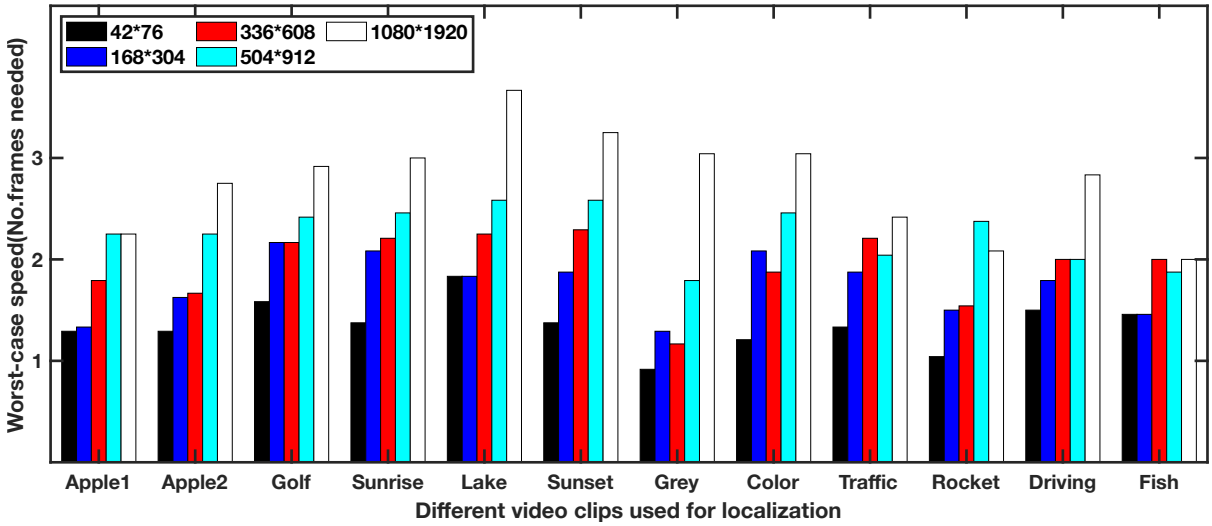

Fig. 30. Evaluating the worst-case localization speed at different video resolutions.

is an iterative bit-by-bit converging (logarithmic) process, whereby the pool of possible locations shrink by a probabilistic constant factor bit-by-bit.

8 DISCUSSIONS

Does BitLight only work with DLP projectors? How about other display devices? The design of BitLight can be only used with display devices with sequential time-division capabilities. DLP projectors are the most suitable hardwares readily available on the market that have this capability inherently built-in. Traditional

display devices such as OLED/LCD displays or traditional LCD projectors are apparently not usable for our design.

How do you compare BitLight with computer vision or other approaches? Admittedly, traditional computer vision based and ultrasonic approaches have achieved very good finger tracking performance even without light sensors. As a result, they are much more convenient in practical use-cases. Understandably, future research needs to overcome the usability inconvenience of BitLight.

How does BitLight perform in noisy environment or without line of sight? Like most visible light-based systems [13, 15, 18, 30, 34], BitLight also requires line of sight to operate. However since BitLight encourages mass deployment of low-cost light sensors as location tags, a number of receivers will perform well even if some others are obstructed. Deliberate disruptions such as blocking the projector light or saturating the photodiode with strong torch light will disrupt the system operation. Common indoor activities such as people walking by or moving furnitures have been verified to not affect system operations.

9 RELATED WORK AND ALTERNATIVE APPROACHES

9.1 Screen-camera Communication

Screen-camera communication systems encode data in the displayed visual images and use cameras to extract the embedded data. Past works have proposed various techniques and coding schemes to improves different aspects of this paradigm. Strata by Wenjun et al. [9] and Bokode by Mohan et al. [19] work on improving long-distance decoding for the traditional bar-code readings; COBRA by Tian et al. [7] and PixNet by Perli et al. [21] aim to boost the data throughput; Lightsync by Wenjun et al. [8] improves the decoding performance of low frame-rate cameras. Many recent researches aim at achieving unobtrusive image encoding that also preserves the human-perceived visual images. Many different frame-level coding techniques have been adopted towards this goal. Hi-light by Tianxing et al. [16] uses alpha channel that embeds bits into pixel translucency changes; Inframe by Anran et al. [32] and Uber-in-Light by Mostafa et al. [10] make complementary changes in the pixel luminance of two consecutive frames and the leverage the spatial-temporal flicker-fusion property of human eyes to ensure minimum visual disturbance. Admittedly, they have achieved the desirable goal of unobtrusive data transmission. Since frame-level encoding carries only 1 bit per pixel per frame, the overall data transfer throughput relies on high image resolution. Their encodings are indifferent to the underlying displayed video and could thus only encode a fixed and limited number of bits of information per pixel location. If K bits (K>1) is desired from such encoding/decoding, the receiver must be able to recognize 2^K unique levels of signals, which are highly unstable. For example, a light sensor that could differentiate four different light intensities could transfer two bits per pixel location per frame.

9.2 Visible-light Positioning Systems

Light positioning systems stem from VLC and have attracted increasing attention recently as the visible light spectrum complements the existing RF spectrum well in indoor space. Existing research studies primarily fall into three categories based on the optical source and receiver types: 1)LED light fixtures with light-sensor receivers, 2)directional optical projectors with light-sensor receivers, and 3)visual tags with camera receivers.

9.2.1 LED Light Fixtures with Light-sensor Receivers. Early LED-based localization systems position the light-sensor receivers to the nearest LED light source with the highest uniquely identifiable received signal strength (RSS) [6, 24, 40] and have achieved only coarse-grained granularity. Epsilon by Liqun et al. [14] further improves the accuracy by using multi-lateration [35, 42] with Lambertian light intensity attenuation model to estimate to distances from the light receiver to multiple optical sources. Prince et al. in [23] makes another improvement to account for the orientation of the light-sensor in using IMU sensor. Pulsar by Chi et al. [41] further proposes to

use dual-PD light sensor to obtain AoA information. Okuli by Chi et al. [39] uses a light propagation/reflection model to uses a finger.

9.2.2 Projector-based VLP with Light-sensor Receivers. Projector-based systems [18, 30, 34] generate a sequence of timed spatio-temporal events of on/off light patterns to localize the light sensors. Directional optical-path geometry is used to calculate the 3D positions [18, 34]. Due to the line-of-sight nature of optical paths, such systems are verified to often have a superior robustness compared with LED-based systems based on an inaccurate signal strength propagation model. BitLight is new type of projector-based VLP system that utilizes the ultra-fine-granularity time-division capability of DLP projectors and ultra-fast sampling capability of photodiode for sensor-display localization.

9.2.3 Visual Tags and Camera Systems. Some camera-based systems calculates locations using angle-of-arrival (AoA) approaches by processing captured images of either uniquely identifiable LED lamps on the ceiling [13, 38] or images of reflected light from passive optical stickers. StarDust by Stoleru et al. [31] is a typical system that localizes optical passive retro-reflectors by capturing its images with cameras. RainbowLight by Lingkun et al. [15] is another innovate approach that uses multiple directional optical stickers, made from special polarizer and birefringence materials, to 3D localize a mobile camera that captures an image of the stickers. Future researches may dig into ways to combine such systems with the time-division capability of DLP projectors, such as using DLP projectors as the light source etc. Camera systems are however considered by many to be intrusive to user privacy.

9.3 Projector-camera Systems

Systems that make use of projectors to generate imperceptible light patterns, which are then captured by cameras have enabled many interesting applications other than sensor-display localization.

9.3.1 3D Avatar Reconstruction. "3D Talking Heads" by Raskar et al. [25] is designed for 3D digital reconstruction of participants at remote sites for teleconferencing. It uses DLP projectors to project structured light patterns of vertical bars, and two cameras to acquire 3D geometry/texture information and subsequently reconstruct the 3D computer graphics model. The structured light pattern is made imperceptible to human eyes by using the frame-level encoding method of two complementary images projected in quick succession, thus appearing as a uniform grey level.

9.3.2 Future Office. "Office of the future" by Raskar et al. [26] is another research effort that conceptualizes the futuristic use-case of replacing all office lights with projectors to obtain precise control over all of the light in the office. Synchronized cameras are imagined to capture the geometry and reflectance information for all visible surfaces in the office, so that one can project images on the surfaces, render images of the surfaces, or interpret changes in the surfaces. In order to generate imperceptible structured light patterns provide lighting at the same time, it uses the frame-level encoding of two complimentary image patterns played in quick succession, integrated by the visual system and appears as unified grayish white lighting.

9.3.3 Steganography. Projectors and cameras are often used for steganography or watermarking applications through imperceptible data embedding. The existing approaches do not make full use of the ultra-fine time-division capabilities of DLP projectors, hence they often use a video-independent image-processing approach of making minimal changes to the original content as much as possible, such as by modifying the image colors by changing the least significant bit of the colors [1, 12, 17, 28, 33]. Future researches could make use of the encoding/decoding approach of BitLight and the fine time-division capabilities of DLP projectors for steganography or even data-communication purposes by coupling it with a high-speed camera.

10 CONCLUSION

BitLight is an endeavor to reuse the visible light medium in a transparent fashion to achieve the additional benefit of localizing light sensors. It makes full use of the ultra-fine time-division capabilities of DLP projectors to transmit a significantly larger data throughput per pixel, which empowers the rapid localization behind the application of sensor-display interaction.

REFERENCES

[1] R Balaji and Garewal Naveen. 2011. Secure data transmission using video Steganography. In *2011 IEEE International Conference on Electro/Information Technology*. IEEE, 1–5.

[2] Horace B Barlow, E Baumgardt, Jacob Nachmias, H Richard Blackwell, E Aulhorn, H Harms, Eric G Heinemann, G Westheimer, Adriana Fiorentini, Robert M Boynton, et al. 1972. Visual psychophysics. (1972).

[3] Thomas Boutell and T Lane. 1997. Png (portable network graphics) specification version 1.0. *Network Working Group* (1997), 1–102.

[4] Charbax. 2017. *Sony Xperia Touch, projector turns any table or wall into a 23" Android tablet.* https://www.youtube.com/watch?v=M1wzB9yaymE

[5] RL De Valois and KK De Valois. 1988. Spatial vision Oxford University Press. *New York* (1988).

[6] Aaron Ganick and Daniel Ryan. 2014. Self identifying modulated light source. US Patent 8,866,391.

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

[8] Wenjun Hu, Hao Gu, and Qifan Pu. 2013. Lightsync: Unsynchronized visual communication over screen-camera links. In *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 15–26.

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

[10] Mostafa Izz, Zhongyuan Li, Hongbo Liu, Yingying Chen, and Feng Li. 2016. Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[11] Michael Kalloniatis and Charles Luu. 2007. Temporal resolution. (2007).

[12] R Kavitha and A Murugan. 2007. Lossless steganography on avi file using swapping algorithm. In *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, Vol. 4. IEEE, 83–88.

[13] Ye-Sheng Kuo, Pat Pannuto, Ko-Jen Hsiao, and Prabal Dutta. 2014. Luxapose: Indoor positioning with mobile phones and visible light. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 447–458.

[14] Liqun Li, Pan Hu, Chunyi Peng, Guobin Shen, and Feng Zhao. 2014. Epsilon: A visible light based positioning system. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. 331–343.

[15] Lingkun Li, Pengjin Xie, and Jiliang Wang. 2018. RainbowLight: Design and Implementation of a Low Cost Ambient Light Positioning System. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 807–809.

[16] Tianxing Li, Chuankai An, Andrew T Campbell, and Xia Zhou. 2015. Hilight: Hiding bits in pixel translucency changes. *ACM SIGMOBILE Mobile Computing and Communications Review* 18, 3 (2015), 62–70.

[17] Xin Liao, Qiao-yan Wen, and Jie Zhang. 2011. A steganographic method for digital images with four-pixel differencing and modified LSB substitution. *Journal of visual communication and image representation* 22, 1 (2011), 1–8.

[18] Song Liu and Tian He. 2017. SmartLight: Light-weight 3D Indoor Localization Using a Single LED Lamp. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 11.

[19] Ankit Mohan, Grace Woo, Shinsaku Hiura, Quinn Smithwick, and Ramesh Raskar. 2009. Bokode: imperceptible visual tags for camera based interaction from a distance. In *ACM Transactions on Graphics (TOG)*, Vol. 28. ACM, 98.

[20] N/A. 2020. *Shutter speed.* https://en.wikipedia.org/wiki/Shutter_speed

[21] Samuel David Perli, Nabeel Ahmed, and Dina Katabi. 2010. PixNet: Interference-free wireless links using LCD-camera pairs. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 137–148.

[22] Charles Poynton. 2012. *Digital video and HD: Algorithms and Interfaces.* Elsevier.

[23] Gregary B Prince and Thomas DC Little. 2012. A two phase hybrid RSS/AoA algorithm for indoor device localization using visible light. In *Global Communications Conference (GLOBECOM), 2012 IEEE*. IEEE, 3347–3352.

[24] Niranjini Rajagopal, Patrick Lazik, and Anthony Rowe. 2014. Visual light landmarks for mobile devices. In *Proceedings of the 13th international symposium on Information processing in sensor networks*. IEEE Press, 249–260.

[25] Ramesh Raskar, Henry Fuchs, Greg Welch, Adam Lake, and Matt Cutts. 1998. 3D Talking heads: Image based modeling at interactive rates using structured light projection. (1998).

[26] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. 179–188.
[27] Paul Read and Mark-Paul Meyer. 2000. *Restoration of motion picture film*. Elsevier.
[28] Toby Sharp. 2001. An implementation of key-based digital signal steganography. In *International Workshop on Information Hiding*. Springer, 13–26.
[29] Ernst Simonson and Josef Brozek. 1952. Flicker fusion frequency: background and applications. *Physiological reviews* 32, 3 (1952), 349–378.
[30] Radu Stoleru, Tian He, John A Stankovic, and David Luebke. 2005. A high-accuracy, low-cost localization system for wireless sensor networks. In *Proceedings of the 3rd international conference on Embedded networked sensor systems*. ACM, 13–26.
[31] Radu Stoleru, Pascal Vicaire, Tian He, and John A Stankovic. 2006. Stardust: A flexible architecture for passive localization in wireless sensor networks. In *Proceedings of the 4th international conference on Embedded networked sensor systems*. ACM, 57–70.
[32] Anran Wang, Zhuoran Li, Chunyi Peng, Guobin Shen, Gan Fang, and Bing Zeng. 2015. Inframe++: Achieve simultaneous screen-human viewing and hidden screen-camera communication. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 181–195.
[33] H-C Wu, N-I Wu, C-S Tsai, and M-S Hwang. 2005. Image steganographic scheme based on pixel-value differencing and LSB replacement methods. *IEE Proceedings-Vision, Image and Signal Processing* 152, 5 (2005), 611–615.
[34] Bo Xie, Guang Tan, and Tian He. 2015. Spinlight: A high accuracy and robust light positioning system for indoor applications. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 211–223.
[35] S-H Yang, E-M Jeong, D-R Kim, H-S Kim, Y-H Son, and S-K Han. 2013. Indoor three-dimensional location estimation based on LED visible light communication. *Electronics Letters* 49, 1 (2013), 54–56.
[36] Yinyu Ye. 1998. On the complexity of approximating a KKT point of quadratic programming. *Mathematical programming* 80, 2 (1998), 195–211.
[37] Yinyu Ye and Edison Tse. 1989. An extension of Karmarkar's projective algorithm for convex quadratic programming. *Mathematical programming* 44, 1-3 (1989), 157–179.
[38] Masaki Yoshino, Shinichiro Haruyama, and Masao Nakagawa. 2008. High-accuracy positioning system using visible LED lights and image sensor. In *Radio and Wireless Symposium, 2008 IEEE*. IEEE, 439–442.
[39] Chi Zhang, Josh Tabor, Jialiang Zhang, and Xinyu Zhang. 2015. Extending mobile interaction through near-field visible light sensing. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 345–357.
[40] Chi Zhang and Xinyu Zhang. 2016. LiTell: robust indoor localization using unmodified light fixtures. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 230–242.
[41] Chi Zhang and Xinyu Zhang. 2017. Pulsar: Towards ubiquitous visible light localization. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 208–221.
[42] Zhou Zhou, Mohsen Kavehrad, and Peng Deng. 2012. Indoor positioning algorithm using light-emitting diode visible light communications. *Optical Engineering* 51, 8 (2012), 085009–1.