Computer Vision Methods for Visual MIMO optical system

Wenjia Yuan Kristin Dana Michael Varga
Ashwin Ashok Marco Gruteser Narayan Mandayam
Rutgers, The State University of New Jersey
wenjiay@eden.rutges.edu, kdana@ece.rutgers.edu, mfvarga@rutgers.edu
{aashok,gruteser,narayan}@winlab.rutgers.edu

Abstract

Cameras have become commonplace in phones, laptops, music-players and handheld games. Similarly, light emitting displays are prevalent in the form of electronic billboards, televisions, computer monitors, and hand-held devices. The prevalence of cameras and displays in our society creates a novel opportunity to build camera-based optical wireless communication systems based on a concept called visual MIMO. We extend the common term MIMO from the field of communications ("multiple-input multiple-output") that is typically used to describe multiple antenna, multiple transmitter radio frequency communications channel. In the visual MIMO communications paradigm, the transmitters are light-emitting devices such as electronic displays and cameras are the receivers. In this paper we discuss and address several challenges in creating a visual MIMO channel. These challenges include: (1) electronic display detection, (2) embedding the transmission signal in the display video, and (3) system characterization for electronic display appearance.

1. Introduction

In recent years, the presence of cameras and light emitting devices has become pervasive in our indoor and outdoor environment. Integrated cameras are prevalent on phones, laptops, e-readers, music-players and many handheld games. Light emitting displays, signage and monitors are prevalent in the form on electronic billboards, computer monitors, information kiosks, mobile displays and in smaller dimensions on hand-held devices. The widespread cameras and displays in our society creates an exciting and novel opportunity to build camera-based optical wireless communication systems based on a concept called *visual MIMO* [3, 2]. In all communications frameworks, the key components are transmitters (e.g. RF, radio frequency) and receivers (antennas). In the the visual MIMO communications paradigm, transmitters are light-emitting devices such as electronic displays and cameras are the receivers. Given the spatial arrangement of pixels on a display and the array of light sensing elements in a camera, the MIMO "multiple-input multiple-output" description is clearly applicable. This paradigm is a unique intersection of the field of computer vision and communications.

As discussed in [3], the approach has several distinct properties that can present advantages over RF based wireless communications. It allows highly directional transmission and reception, rendering it virtually interference-free and attractive for very dense congested environments. Also, such transmissions are hard to detect and intercept, which is beneficial for security applications. Furthermore, visual MIMO is a low-cost alternative to RF because it takes advantage of existing cameras and electronic displays.

Diverse applications of technology using cameras and displays for communications are identifiable especially if the transmission signals are embedded in the existing display image or video. This signal embedding enables dual use of electronic displays so that visual observation for human observers coexists with a visual MIMO wireless communications channel. This approach would enable novel advertising applications such as smartphone users pointing cell phone cameras at an electronic billboards to receive further information such as a purchase URL. Also, consider applications where exhibit information from a kiosk display is transferred to cameras on ipods in order to obtain customized audio museum tours. Localization in the event of an emergency is another potential application. A cell phone display held up to an existing surveillance camera/receiver could transmit the precise location of a person within a high-rise building to assist emergency first-responders.

In this paper we discuss and address several computer vision challenges in creating a visual MIMO channel. These challenges include: (1) electronic display detection, (2) signal embedding, (3) characterization of electronic display appearance. Electronic display detection is an important challenge because the camera or display is mobile and the scene is dynamic in real world applications. When communicating with RF, signals are detected at particular frequencies in the transmission bandwidth. But with a visual channel, finding the transmitter in the image of the scene requires the type of processing that is the domain of computer vision: recognizing a particular object or pattern in a scene. We present methods for detecting the transmission pattern (i.e. randomized checkerboard) under changes in camera pose. When the transmission pattern is to go unnoticed, i.e. when the monitor is to be used for signal transmission as well as for displaying an unrelated video of image to a human observer, signal embedding is needed. Signal embedding is close to the area of watermarking and steganography. However, we have the additional challenge that the digital image is transmitted on an electronic display and then observed by a camera. Therefore the digital signal is converted to the analog signal (visible light) and undergoes scene-dependent photometric and geometric transformations before being observed by the camera and converted back to a digital signal. As such, simple traditional methods for signal embedding are not applicable. In this paper we discuss an intensity modulation method to handle signal embedding in a visual MIMO channel. Another interesting issue to consider is the appearance of the transmission pattern in the electronic display as a function of camera pose. In this paper, we consider the case of an LCD electronic display and measure its appearance change with viewing angle. While the computer vision literature has concentrated on reflective objects, a light-emitting display cannot be characterized (even approximately) by lambertian and specular shading models. The angular dependence of light intensity is fundamentally different. Our measurements provide a useful system characterization metric by showing the strong dependence of LCD light intensity with viewing angle. This quantification will be essential for determining thresholds for identifying the on-off state in the transmission pattern in order to interpret the transmission at the receiver. Finally, we perform an initial system test by sending signals using an LCD monitor, detecting the transmission pattern and computing the error rate as a function of camera pose. Our initial results indicate the viability of the visual MIMO framework.

2. Related work

The concept of using cameras as receivers in a communication paradigm is novel and the literature on the topic is sparse. In our prior work [3, 2], we have characterized the potential capacity and effective bandwidth of this system with application in inter-vehicle communication (cameras in cars receiving signals from LED's in taillights). Other related work are for inter-vehicle communications [11] and traffic light to vehicle communications [13]. Other work has investigated channel modeling [9] and multiplexing [1]. More recently, researchers of the MIT Bokode project [10] have applied computational photography to camera based communications. The goals of this work are fundamentally interdisciplinary and visual MIMO requires rethinking the physical layer when compared to traditional baseband signal processing. In the computer vision community, the concept of using vision algorithms in the physical layer of a communications channel is novel.

3. Electronic Display Detection

In this work, we assume that the transmission pattern is an array of ones and zeros transmitted as black and white squares with the appearance of a randomized checkerboard to represent the transmitted data. There are two important steps in detecting the pattern. First, the checkerboard region is localized from the received images using the cues of lines and corners. Second, each detected block with the pattern must be assigned with a black or white label (Section 3.2). For locating the region two novel methods are proposed in Section 3.1. One method is based on a single image, and the other uses two temporally sequential frames. An alternative method using polarization is also presented in Section 3.3 that takes advantage of an LCD monitor's polarization property and is relevant when specialized cameras with polarization filters can be employed.

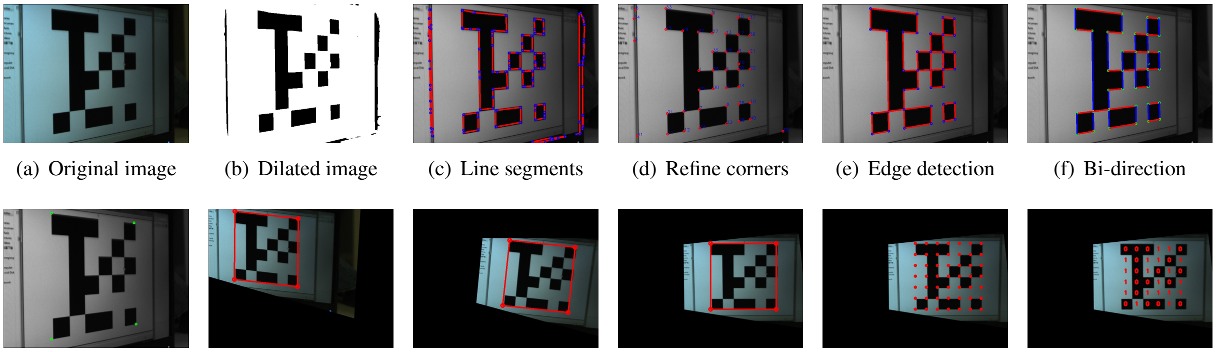
3.1. Random Checkerboard Localization

3.1.1 Single Frame Detection

Lines and corners are key cues to detect random checkerboards. Our goal is to obtain four borders of the random checkerboard or directly detect the four corners. However, unlike a true checkerboard pattern with pre-defined variation, in the random pattern of the transmitted message, for example, if many neighboring blocks are of the same intensity, fewer edges exist for block localization. To circumvent this situation, we fix four known intensity anchor blocks at the four corners as shown in Figure 1(a).

The detection procedure is listed in Table 1. In order to remove noise from the environment, we dilate the binary image as in Figure 1(b). Then the Hough transformation is used to detect line segments (Figure 1(c)). The ends of these line segments are corner candidates, which encode both line and corner information. The corners from the dilated image are then refined to the sub-pixel level. Harris corner detection further selects good corners from these refined candidates (Figure 1(d)). The results of processing one frame are illustrated in Figure 1.

Line segments are detected by connecting any pair of these good corners (Figure 1(e)). The vertical and horizontal lines could be detected based on the orientation of line segments. In Figure 1(f), the horizontal lines are in red and the vertical ones are in blue. Based on four outermost borders from two directions, four outermost corners of the


(a) Original image (b) Dilated image (c) Line segments (d) Refine corners (e) Edge detection (f) Bi-direction

(g) Four corners (h) Affine rectification (i) Metric rectification (j) Angle adjustment (k) Localize blocks (l) Assign labels

Figure 1. The procedure and experiment results for single-frame random checkerboard detection. Anchor blocks are all set as black, locating at four corners. The random checkerboard is displayed on an LCD monitor.

Input: the processed frame i_k, the number of blocks in rows and columns: $nBlkPerRow$ and $nBlkPerCol$.

Output: the black or white assignment for each block on i_k.

1. Dilate the binary image of i_k.
2. Obtain candidate corners by line segment detection algorithm based on Hough transformation.
3. Select good corners with Harris corner detection method.
4. Find vertical and horizontal line segments based on the corners selected from the last step.
5. Find four outermost corners of the random checkerboard.
6. Make affine/metric rectification and rotation adjustment.
7. Localize all blocks from the random checkerboard.
8. Block intensity detection.

Table 1. **Method I:** single-frame random checkerboard detection. In this algorithm, four black corner blocks are anchors. Metric rectification is used for the random checkerboard with square blocks.

random checkerboard can be localized (Figure 1(g)).

With four outermost corners of the random checkerboard and two parameters $nBlkPerRow$, $nBlkPerCol$ (i.e. the number of blocks in rows in columns), we can identify the location of each block. We know each block is square, so the perspective distortion can be overcome by affine rectification (Figure 1(h)) and metric rectification (Figure 1(i)) [6]. After rotation adjustment (Figure 1(j)), locations of blocks can be identified in the rectified image (Figure 1(k)). Then we may classify each block as black or white corresponding to a transmission of '0' or '1'. Because of the intensity dependence with angle in LCD monitors (Section 5.1) and surface reflectance, the intensity of white and black from the received image is not an ideal 255 and 0. A discussion on the classification of a block as black or white is given in Section 3.2.

3.1.2 Two-frame Method

Differences between two temporally sequential image frames are useful in localizing the transmission pattern (randomized checkerboard). To ensure the identification of four outermost corners, we set four anchor blocks at corners of the random checkerboard. The anchor blocks are set in a slightly different way for the two-frame method. All anchor blocks from odd temporal index frames are black, and from all even-index frames are white.

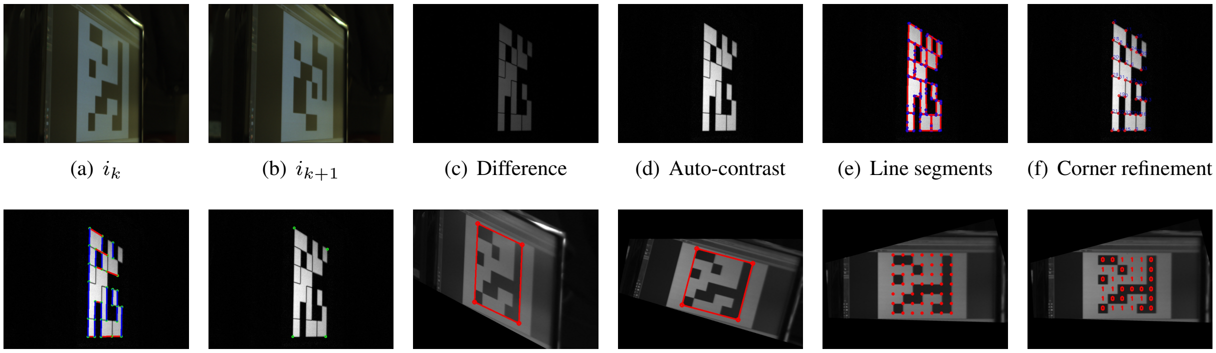
The procedure is listed in Table 2 and the processing results of the k^{th} frame are illustrated in Figure 2. The frame i_{k+1} (Figure 2(b)) serves as a reference frame for i_k. We get the absolute difference image between i_k and i_{k+1} in Figure 2(c). Auto-contrast is applied (Figure 2(d)) and then the method proceeds as in the single frame method of Section 3.1.1: get the line segments (Figure 2(e)), refine corners with Harris corner detection (Figure 2(f)), take vertical and horizontal line segments (Figure 2(g)), and finally localize the random checkerboard by getting four outermost corners (Figure 2(h)). The image is then adjusted with affine rectification (Figure 2(i)) and metric rectification (Figure 2(j)) [6]. The final detection result is illustrated in Figure 2(l).

Compared with the method based on a single frame in Section 3.1.1, the two-frame method is more robust. The price paid is the cost of frame-to-frame alignment that is needed when the camera and/or display is mobile.

3.2. Black and White Classification

After localizing the transmission pattern, the next step is to assign black or white label to each block. However, the intensity of received images changes with viewing angles (as discussed further in Section 5.1).

To show the change in intensity with angles, we select nine viewing angles ranging from 0 to 80 degrees and six out of these nine angles are listed in Figure 8. The definition of the viewing angle is illustrated in Figure 7. We can



(a) i_k (b) i_{k+1} (c) Difference (d) Auto-contrast (e) Line segments (f) Corner refinement

(g) Bi-direction (h) Four corners (i) Affine rectification (j) Metric rectification (k) Locations of blocks (l) Label assignments

Figure 2. The procedure and experiment results with random checkerboard detection based on two temporally sequential frames displayed on an LCD monitor. All the anchor blocks are at four corners. From odd temporal index frames like Figure 2(a), all anchor blocks are black, while they are all white from odd temporal index frames like Figure 2(b).

observe that larger viewing angles correspond the darker screens. More specific results are presented in Figure 3.

For interpreting the transmission pattern, we use average intensity to represent each block. For each angle, the average intensities of all blocks from the first 20 frames are collected. For the white blocks, we compute the mean, minimal and maximal intensity and represent the results with a solid line in Figure 3: points indicate means of all white blocks at a particular angle; ranges are labeled between minimal and maximal values. Similarly, we describe the results for black blocks with a dash-dotted line.

We observe that the average intensity for white blocks changes greatly with angles. When angles close to 90 degree, i.e. the screen plane is near parallel to the line between the camera and the screen center, the intensities of black and white blocks approach to each other. However, their intensities are still separable. We employ the unsupervised clustering method agglomerative hierarchical clustering [5] to group the average intensities for all blocks at certain angle into two groups labeled as white and black.

Input: two neighboring frames i_k and i_{k+1}, number of blocks in rows and columns $nBlkPerRow$, and $nBlkPerCol$.

Output: the black or white assignment for each block on i_k.

1. Get the absolute difference between i_k and i_{k+1}.
2. Rescale the difference with auto-contrast algorithm.
3. Obtain candidate corners by line segment detection algorithm based on Hough transformation.
4. Select good corners with Harris corner detection method.
5. Find vertical and horizontal line segments based on the corners selected from the last step.
6. Find four outermost corners of the random checkerboard.
7. Make affine/metric rectification and rotation adjustment.
8. Localize all blocks.
9. Block intensity detection.

Table 2. **Method II:** random checkerboard detection with difference between two neighboring frames. In this algorithm, four corner blocks are set as anchors. From all the odd frames, anchor blocks are black; while from even frames, they are white. Metric rectification is used for the random checkerboard with square blocks.

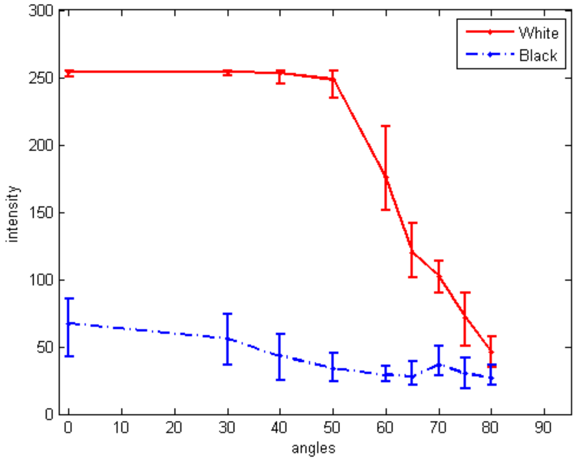


Figure 3. Changes of black/white intensities with viewing angles.

3.3. Polarization-based Detection

Another method for detecting the electronic display does not use any property of the transmission pattern. Instead, the polarization property of the electronic display is used. We assume that the light from the display is polarized (such as an LCD display). For example, most computer monitors and cell phone displays are polarized. For visual MIMO, the transmission pattern is displayed on the electronic display or the embedded signal is displayed on the display. The camera views both the electronic display and the background. However, light from the background is not typically polarized. Therefore, to detect the display within the scene image, the light modulation that occurs with a rotating polarizer can be employed. Specifically, as shown in Figure 4, pixels in the LCD display region will undergo a change in intensity when the polarizer is rotated or when the camera with a lens-mounted polarizer is rotated. This intensity change is sinusoidal and can typically be detected using two or more orientations of the polarizer or camera.

4. Signal Embedding

In order to simultaneously use a display for its original purpose and as a communication channel, we need the ability to embed the transmission signal in the display video. This goal is related to the field of steganography. However unlike steganography or watermarking in the digital domain [8, 7], our camera-display system displays and then observes the image. This pipeline includes a conversion from a digital signal $i[x, y]$ to the analog signal $i(x, y)$ representing the light transmitted by the display. The camera observing the display has a light sensing array (e.g. CCD) that converts the received signal and creates the digital signal $\tilde{i}$. Clearly, $i[x, y]$ is not equal to $\tilde{i}[x, y]$ because of the photometric and geometric transformations in image formation. Consequently, some standard techniques in steganography will not work for signal embedding in this paradigm. For example, LSB embedding uses the least significant bit for signal hiding. This approach clearly would not work because of the differences in the received and transmitted signal. We use the term *photographic steganography* for the approach of hiding signals in observed imagery to distinguish it from methods where the image to be transmitted and received are both digital signals.

We present a novel approach for photographic steganography that employs intensity modulation. The concept of intensity modulation has had various applications in computer vision including illumination multiplexing [12] and polarization multiplexing [4] for parsing the single-illuminant from multi-illuminant images. In the application of photographic steganography, intensity modulation can be used to hide and embed a signal s by transmitting two perturbed signals i_1 and i_2 that are defined as follows

$$i_1 = i - \alpha s, \tag{1}$$

$$i_2 = i + \alpha s, \tag{2}$$

which can be expressed in terms of a modulation matrix M:

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = M \begin{bmatrix} i \\ s \end{bmatrix}. \tag{3}$$

The modulation maxtrix M is given by

$$M = \begin{bmatrix} 1 & \alpha \\ 1 & -\alpha \end{bmatrix}. \tag{4}$$

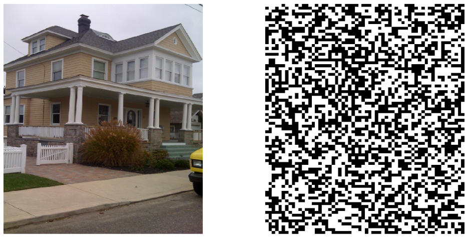


Figure 5. The original signal i (left) and the signal to be transmitted s (right). Here s is a binary signal corresponding to on-off keying in a communications channel.

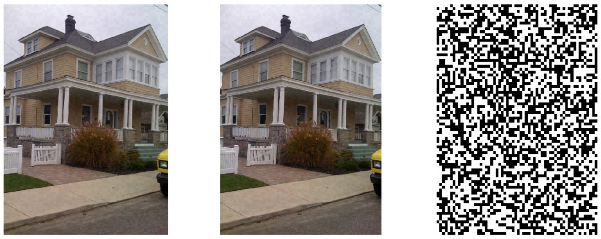


Figure 6. Separation of the embedded signal. Two frames used for embedding the signal i_1 (left) and i_2 (right) as given by Equation 3. The separated signal s (right) as given by Equation 5.

Note that x, y dependence for i, s, i_1, and i_2 has been omitted for notational simplicity. Since M is invertible for $\alpha \neq 0$, the original signal i and the transmission signal s can be obtained with

$$\begin{bmatrix} i \\ s \end{bmatrix} = M^{-1} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix}. \tag{5}$$

The frame rate is reduced by a factor of 2 in this approach. We assume that a higher frame rate camera can be employed so that recording at twice the desired transmission rate can be achieved. Additionally, we assume that an image stabilization algorithm can be employed to account for frame-to-frame camera motion. Demonstration of this approach is illustrated in Figure 5 and Figure 6.

5. Electronic Display Appearance

Characterization of the the electronic display appearance largely influences the accuracy of the data received by the camera. In this section, we investigate three factors: angle, distance, and block size in pixels. In the experiment, we use a transmission pattern (random checkerboard) video with 200 frames. For each frame, the signal is transmitted from an LCD screen to a camera. After comparing the recovered and the transmitted data, we compute the error rate by a ratio by dividing the number of incorrectly-detected blocks by the total number of the blocks in the video.

5.1. Error rate by angles

Light emitting devices do not radiate light in all directions with the same power as is illustrated in Figure 3 for the

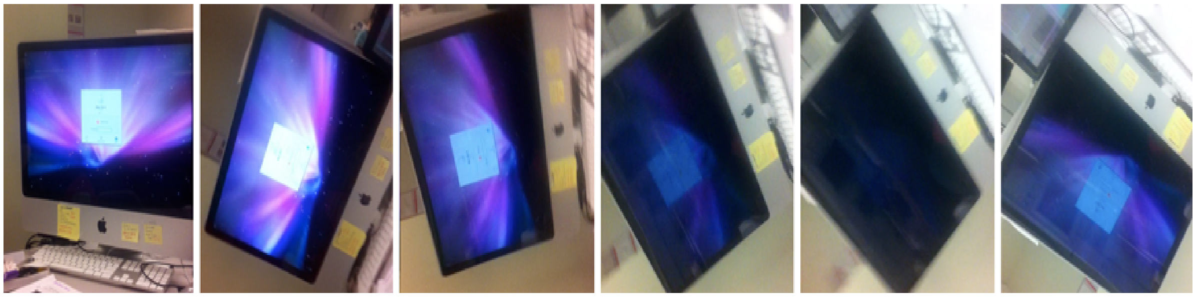


Figure 4. The polarized property of screens: screens emit light in a polarized beam way. A grid-wire polarized sheet is attached to the camera. When the orientation of the emitting beam is vertical to that of the wire from the polarizer, no light goes through the polarized sheet and the screen turns to be dark. This property of polarization can be used for detecting the screen region.

LCD monitor. When the angle increases, less light is emitted per unit area and the received image darkens. The intensity of the white blocks changes significantly more than the black intensity. When the viewing angle defined in Figure 7 is close to 90 degree, it becomes difficult to discern between black and white.

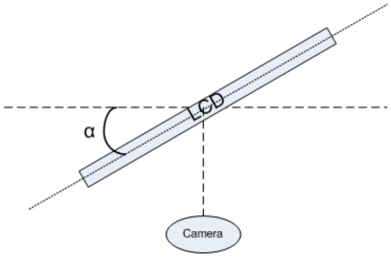
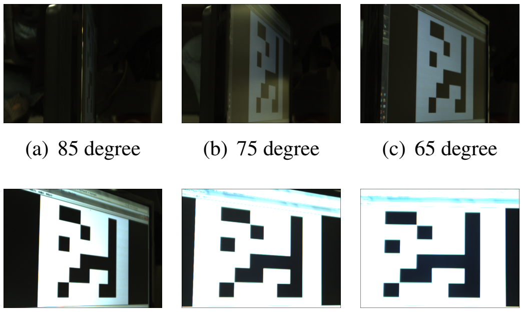


Figure 7. The model to test white and black intensities changing with angle α. When $\alpha = 0$, the LCD plane faces directly to the camera.



(a) 85 degree (b) 75 degree (c) 65 degree

(d) 50 degree (e) 30 degree (f) 0 degree

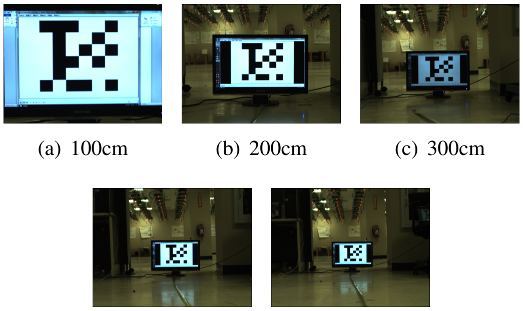
Figure 8. An LCD monitor for displaying random checkerboards with different angles.

angles(°)	0	30	40	50	60	70	75
error rate (%)	0	0	0	0	0	29.33	44.47

Table 3. Error rate over angles. The angle is denoted as α in Figure 7. When $\alpha = 0$, the LCD plane faces directly to the camera. When α increases, the system suffers more perspective distortion and pixel blurring and the system has a higher error rate.

The measured error rates in angles are listed in Table 3. We can observe that error rates zero from 0 to 60 degrees, indicating the black/white classification was successful. The corresponding images are shown in Figure 8 and it's clear that there is a large enough contrast to label the white and black blocks. When the angle α increases past $60°$, the contrast decreases and corners are smoothed by the effect of lens blurring. As expected, this leads to correspondingly larger error rates.

5.2. Error rate by distances



(a) 100cm (b) 200cm (c) 300cm

(d) 400cm (e) 500cm

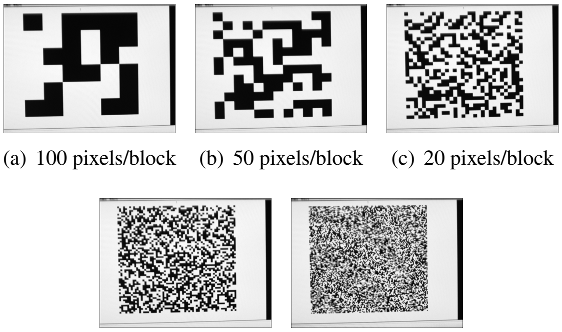
Figure 9. An LCD monitor for displaying random checkerboards with different distances.

distances(cm)	100	200	300	400	500
error rate (%)	0	0	0	0	1.15

Table 4. Error rate in distances between the camera and LCD. The system suffers more blurring effects in the detected random checkerboard, because of larger distances. Thus the error rate increases when the distance turns larger. In this experiment, no effect of angles or block sizes is considered.

In this experiment, the distance from camera to monitor is measured in centimeters. Five values of distances are selected as in Figure 9. From Table 4, we can observe that the error rates remains at zero from 100 to 400 centimeters. When the distance increases to 500 centimeters, the error rate increases to 1.15%. As expected, a further distance results in degradation in the ability to spatially resolve the blocks due to blurring and perspective effects.

5.3. Error rate by block sizes



(a) 100 pixels/block (b) 50 pixels/block (c) 20 pixels/block

(d) 10 pixels/block (e) 5 pixels/block

Figure 10. An LCD monitor for displaying random checkerboards with different number of pixels per block.

block sizes(pixels)	5	10	20	50	100
error rate (%)	54.15	20.36	5.47	0.42	0

Table 5. Error rate as a function of block size from the random checkerboard. The block size is the number of pixels in rows/columns of a square block. When block size is smaller, blurring is more apparent and the error rate goes higher.

Higher data rate with a fixed bandwidth is a desired target in communication system design. Within limits, data rate can be increased in a visual MIMO system by decreasing the block size. When the block size approaches the size of a pixel, the difficulty in detecting neighboring black of white pixels increases. We investigate the error rate as a function of block size using square pixels where the size indicates either the width or height of the block in pixels. We select five size values as shown in Figure 10(e). The results are listed in Table 5. When the sizes of blocks are reduced, the corresponding error rate goes up. As expected, we observe that data rate can be improved by decreasing block size. For our system, a block size less than 20 pixels has an associated error rate higher than a 5%.

6. Conclusion and Summary

Visual MIMO communication with cameras and displays provides an interdisciplinary challenge for the fields of wireless communications and computer vision. We have approached the problem from a computer vision perspective and addressed several key issues in developing these methods. We have proposed algorithms and presented initial results for : (1) electronic display detection (2) signal embedding, (3) characterization of electronic display appearance. Future work includes building a robust communications link between display and camera for various messaging tasks in real world scenes.

References

[1] S. Arai, S. Mase, T. Yamazato, T. Endo, T. Fujii, M. Tanimoto, K. Kidono, Y. Kimura, and Y. Ninomiya. Experimental on hierarchical transmission scheme for visible light communication using led traffic light and high-speed camera. *IEEE Vehicular Technology Conference*, 2007. 2

[2] A. Ashok, M. Gruteser, N. Mandayam, T. Kwonz, W. Yuan, M. Vargay, and K. Dana. Rate adaptation in visual mimo. *Communications Society Conference on Sensor, Mesh, and Ad Hoc Communications and Networks*, 2011. 1, 2

[3] A. Ashok, M. Gruteser, N. Mandayam, J. Silva, M. Varga, and K. Dana. Challenge: mobile optical networks through visual mimo. *International conference on Mobile computing and networking*, 2010. 1, 2

[4] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang. Polarization multiplexing and demultiplexing for appearance-based modeling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, (2):362–367, 2007. 5

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, second edition, 2000. 4

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, pages 49–57. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 3

[7] N. F. Johnson, Z. Duric, and S. Jajodia. *Information Hiding: Steganography and Watermarking - Attacks and Countermeasures*. Springer, first edition, 2000. 5

[8] N. F. Johnson and S. Jajodia. Exploring steganography: Seeing the unseen. *IEEE institute of electrical and electronics*, (2):26–35, 1998. 5

[9] T. Komine and M. Nakagawa. Fundamental analysis for visible-light communication system using led lights. *IEEE Transactions on Consumer Electronics*, 2004. 2

[10] S. D. Perli, N. Ahmed, and D. Katabi. Pixnet:designing interference-free wireless links using lcd-camera pairs. *Proceedings of ACM International Conference on Mobile Computing and Networking*, 2010. 2

[11] T. Saito, S. Haruyama, and M. Nakagawa. Inter-vehicle communication and ranging method using led rear lights. *IEEE Communication Systems and Networks*, pages 278–283, 2006. 2

[12] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur. Multiplexing for optimal lighting. *IEEE Trans. Pattern Analysis and Machine Intelligence*, (8):1339–1354, 2007. 5

[13] H. B. C. Wook, S. Haruyama, and M. Nakagawa. Visible light communication with led traffic lights using 2-dimensional image sensor. *Transactions on Fundamentals of Electronics, Communications and*, pages 278–283, 2006. 2