

Dynamic and Invisible Messaging for Visual MIMO

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

Abstract

The growing ubiquity of cameras in hand-held devices and the prevalence of electronic displays in signage creates a novel framework for wireless communications. Traditionally, the term MIMO is used for multiple-input multiple-output where the multiple-input component is a set of radio transmitters and the multiple-output component is a set of radio receivers. We employ the concept of visual MIMO where pixels are transmitters and cameras are receivers. In this manner, the techniques of computer vision can be combined with principles from wireless communications to create an optical line-of-sight communications channel. Two major challenges are addressed: (1) The message for transmission must be embedded in the observed display so that the message is hidden from the observer and the electronic display can simultaneously be used for its originally intended purpose (e.g. signage, advertisements, maps); (2) Photometric and geometric distortions during the imaging process corrupt the information channel between the transmitter display and the receiver camera. These distortions must be modeled and removed. In this paper, we present a real-time messaging paradigm and its implementation in an operational visual MIMO optical systems. As part of the system, we develop a novel algorithm for photographic message extraction which includes automatic display detection, message embedding and message retrieval. Experiments show that the system achieves an average accuracy of 94.6% at the bitrate of 6222.2 bps.

1. Introduction

The growing ubiquity of cameras in hand-held devices and the prevalence of electronic displays in signage and monitors creates a novel framework for wireless communications. In the field of wireless communications, the term MIMO is used for *multiple-input multiple-output* communications. Classically, the multiple-input component is a set of radio transmitters and the multiple-output component is

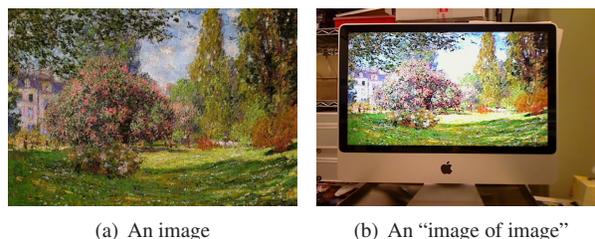


Figure 1. Comparison of “an image” and “an image of image”. Unlike traditional vision problems for images, visual MIMO processes an “image of image”, which undergoes photometric and geometric distortion. Consequently, we use the term *photographic message extraction* as the method to embed and recover hidden messages in a visual MIMO system.

a set of radio receivers. We employ the concept of *visual MIMO* where pixels are transmitters and cameras are receivers. In this manner, the techniques of computer vision can be combined with principles from wireless communications to create an optical line-of-sight communications channel. This concept was introduced in [3, 2, 24] and in this paper we present a working prototype system that enables a hand-held moving camera pointed at a display (such as an electronic billboard, kiosk or computer monitor) to receive a hidden message while a human observer simultaneously sees the display of an unrelated image or textual information.

This computer-vision based communications channel has numerous applications in situations where line-of-sight communications is desirable. For example, the approach enables novel advertising applications such as smartphone users pointing cell phone cameras at electronic billboards to receive further information including documents, movie clips, or a website URL. Another example is a museum application where a kiosk display transfers exhibit information to cell phone camera in order to obtain maps, images and customized audio museum tours. Applications are not limited to hand-held cameras and electronic displays but also include vehicle-vehicle communication, robot-to-robot

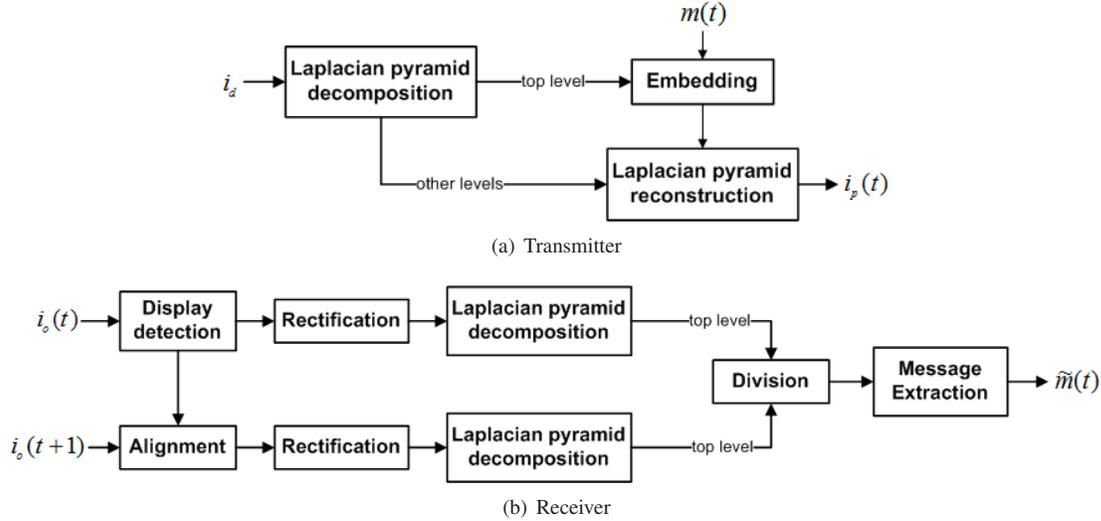


Figure 2. Flow chart of photographic message extraction system for visual MIMO

communication, and hand-held displays to fixed surveillance cameras.

While traditional computer vision concentrates on images of objects, people or scenes; visual MIMO is confronted with the problem of an “image of an image” as illustrated in Figure 1. More specifically, the electronic display is observed by a camera and major challenges must be addressed. First, the message for transmission must be embedded in the observed display so that the message is hidden from the observer and the electronic display can be used for its originally intended purpose (e.g. signage, advertisements, maps). The method of embedding must be such that the transmitted signal is essentially invisible to a human observer of the display. However, the embedding should enable accurate signal recovery at the receiver.

The second major challenge is that photometric and geometric distortions during the imaging process corrupt the information channel between the transmitter display and the receiver camera. These distortions must be modeled and removed. The photometry of image formation determines how the measured pixel values are related to physical properties of the scene such as surface reflectance and the color and brightness of the light source. For a visual MIMO system, a digital image appears on the electronic display and the resulting analog light field is observed by the camera. The transfer function of both the display and the camera must be considered to model the conversion of a transmitted image to a received image. Geometric distortions are caused by the position and orientation of the camera in motion. The camera pose changes in a hand-held application and time-varying distortions from perspective projection are introduced due to the relative location between display and camera. Furthermore, the portion of the image containing

the display must be segmented from the unknown background for signal recovery.

In this paper, we present a real-time visual MIMO optical systems. In building the system we introduce a method of photographic message extraction that includes algorithms for automatic display detection and segmentation, message embedding to hide the transmitted message within the existing display while enabling recovery at the receiver. Experiments show that the system achieves an average accuracy of 94.6% (5.4% bit error rate) at the decoder bitrate of 6222.2 bps. This bitrate is suitable for transmission of URL’s and text information

1.1. Related work

Visual MIMO and related communication theory was introduced in our prior work [3, 2, 24]. An application to vehicle-to-vehicle communication using LED arrays is presented in [22]. The concept of visual MIMO is related to popular QR-codes that can be read with hand-held cameras [10]. However, in the QR-code application the message is fixed and visible. For visual MIMO, the message is dynamic and invisible, enabling the technology to be suitable for communications.

LCD-camera communications is presented in [18] with a time-varying message, but the camera is in a fixed position with respect to the display. Consequently, the computer vision component was not included in this work and the electronic display is not detected, tracked or segmented from the background.

Researchers on the Bokode project [15] and [11] presented systems using invisible message, however that message is a fixed symbol, not a time-varying message. Similarly, traditional watermark approaches typically contained

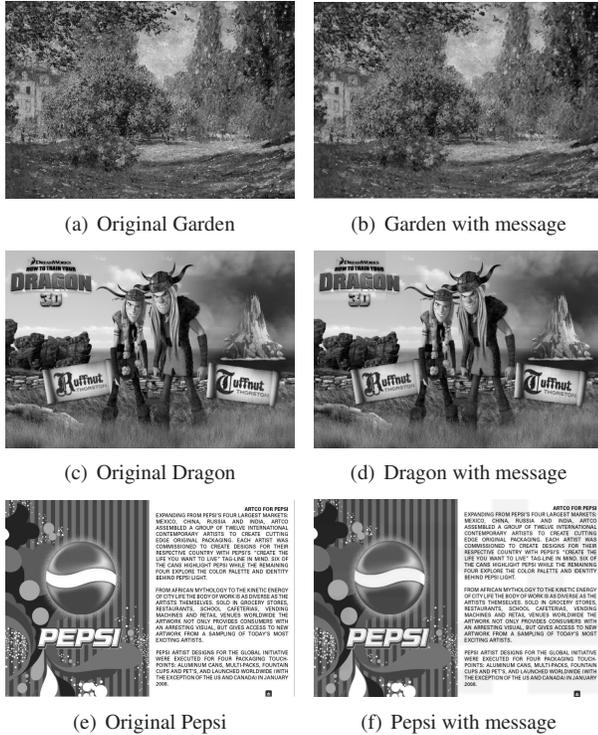


Figure 3. **Comparison of images before and after message embedding.** Left column: the original display images i_d . Right column: the display images $i_p(t)$ with message $m(t)$.

fixed messages [20, 17, 14] applied to digital images. But in our method, when passing through visual MIMO channel, digital images are changed into analog form (via electronic display) and then converted back to digital observed images at the camera. This imaging process creates both photometric and geometric distortions in the observed image. Many robust invariant watermark methods are designed for geometric distortions including invariant moments [1, 7, 12], normalization methods [6], fractal methods [19] and invariant features [21, 23]. Our method handles both geometric and photometric distortions by applying appropriate computer vision methods. Also, our goal is to construct a real-time information transmission system, and the typical goal for digital watermarking is copyright protection.

To our knowledge, our work is the first system that transmits both a time-varying and hidden signal to achieve a camera-display communications system suitable for numerous real-world applications.

2. Methods

We consider an electronic display that is to project the display image i_d . We assume that the display image is static such as an advertisement or decorative display. The problem is to embed a time varying message image $m(t)$ into the

display image as illustrated in Figure 4. The projected image $i_p(t)$ is the display image with the hidden message $m(t)$ embedded. To the human observer $i_p(t)$ should look identical or very similar to i_d and more examples are illustrated in Figure 3. A camera observes the electronic display and images both the display and the background. The position and orientation of the camera with respect to the display is unknown and not fixed. The video frame $i_o(t)$ observed by the camera is a photograph $i_o(t)$. In this sense, the method is photographic message extraction.

Our approach consists of two main components. The first component is the detection and segmentation of the electronic display region from the observed image $i_o(t)$. The second component is the embedding and retrieval method that enables the message $m(t)$ to be hidden in the projected image $i_p(t)$ and recovered from the observed image $i_o(t)$. The framework for our proposed photographic message extraction is illustrated in Figure 2.

2.1. Message Embedding

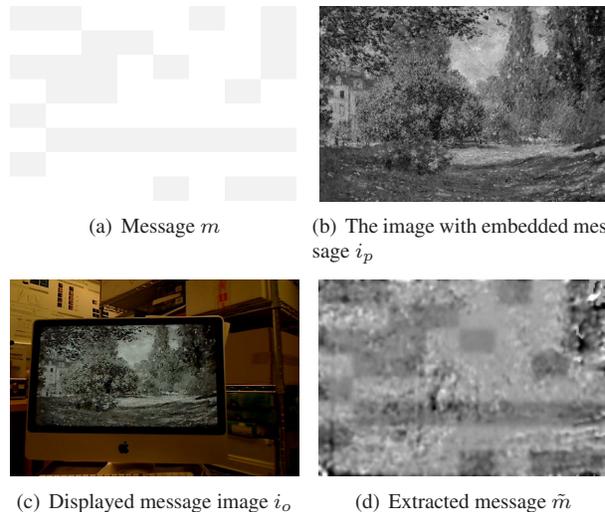


Figure 4. Message embedding with Laplacian pyramid

To facilitate message recovery, a message is embedded in every other frame in the following manner. The projected image for the frame at t_0 is identical to i_d , i.e. $i_p(t_0) = i_d$. The projected image for the frame at t_1 is embedded with the message $m(t_1)$. As such, the message frame rate is half that of the observation camera frame rate. Similarly, $i_p(t_2) = i_d$ and $i_p(t_3)$ contains message $m(t_3)$.

The image with and without the embedded message are two neighboring frames in the displayed sequence. The message embedding is done by using a multiplicative constant on a low pass component of the image. To retrieve that constant the low-frequency components of the $i_o(t_1)$ is divided by the same component of $i_o(t_0)$ (in the segmented

region of the electronic display). The division only works if there is a subpixel alignment between adjacent frames of the observed video sequence $i_o(t)$. Therefore, for each frame the display region must be detected and a frame-to-frame precise alignment is performed (see Section 2.2).

The message structure is a rectangular $N \times M$ grid where each grid sub-rectangle has one bit of information. That is, $m(t)$ for a particular t is a grid-like image of on-off bits as illustrated in Figure 4(a). Unlike traditional watermarking methods, the message image is continually changing according to the bits in the transmitted message. To make $m(t)$ invisible in the projected image, the “off” blocks are set with a fractional value slightly less than 1 (typically 0.9) and the “on” blocks are set to 1. The resulting pattern is blurred by a low-pass filters Gaussian filter so that the spatial transitions are softened and less detectable by the human eye. The motivation here is to change i_d enough to be detectable computationally but not perceptually. The choice of a multiplicative constant as opposed to an additive constant is intentional and well-justified by the fact that this enables a robustness to lighting changes between neighboring frames as inspired by photometric ratios [16].

The choice of embedding the signal in the low frequency component of the original image helps to hide the image from the human observer. High spatial frequency changes such as step-edge changes in intensity are easily detectable by the human visual system. Additionally, distortion of images through observation is often most prominent in the high frequency regions. Consider for example observing an image at a distance, the image area is projected to a smaller region of the imaging sensor and that subsampling combined with lens blur acts as a low pass filter. Classic watermarking embeds messages into the low-frequency region, such as the classic DCT watermarking scheme [5]. For photographic message extraction, the DCT coefficients are not a convenient way to embed the message image $m(t)$. Instead we extract the low frequency component of the image using pyramid processing [4]. The top level of a Laplacian pyramid is a low pass filtered and subsampled version of the original. We define the grid as the same pixel dimensions as the chosen level for embedding.

The flowchart is illustrated in Figure 2. At the transmitter, we decompose the display image i_d into a Laplacian pyramid. The top level stores the most low-frequency information, while the bottom level has comparatively high-frequency information like edges. On the receiver (camera) side, each image is decomposed with a Laplacian pyramid. Recall that two neighboring frames are images with and without the embedded message. We extract $m(t)$ by dividing the top level of the Laplacian pyramid of $i_o(t+1)$ with the top level of the pyramid from $i_o(t)$. The result is illustrated in Figure 4(d). We observe that the message pattern is sufficiently clear for classification.

Input: the first frame $i_o(1)$ and the twentieth frame $i_o(20)$

Output: four corners of the electronic display.

1. Feature extraction for the input images.
2. Obtain the matching features from the same plane by RANSAC homography.
3. Line detection by Hough line detection method.
4. Line refinement and recognition.
5. Final line detection result with optimal ratio method.

Table 1. **Automatic display detection algorithm**

Before the division of the low-frequency components of neighboring frames, a warp of the image to the a frontal view is required so that the predefined grid pattern can be overlaid on the image. In this manner, the grid pattern delineates the individual subrectangles and bits can be read. In order to read the bit patterns, a trained classifier is employed. We assume that the first column of the grid pattern is a fixed and known on/off pattern so that the intensity in these bits can be used to train a classifier in order to determine threshold of average intensity to classify the on/off bits.

An example of the embedding process is illustrated in Figure 4(b). The resultant image $i_p(t)$ is displayed on the monitor, transmitted in the visual MIMO channel by camera capture at the receiver to arrive at the observed image $i_o(t)$ as shown in Figure 4(c).

2.2. Display Detection and Segmentation

The system is comprised of a hand-held moving camera pointed at an electronic display. There are two assumptions: the electronic display is planar and the camera is not fixed. For the display detection and segmentation, we use two close but non-sequential frames, for example, the first frame $i_o(1)$ and the twentieth frame $i_o(20)$ from the video sequence observed by the camera. The goal is to identify the location of the four corners of the display in the image. The algorithm is summarized in Table 1 and the results are illustrated in Figure 5.

The first step is the extraction of Harris corners from the two input images. Then RANSAC homography [8] is used to obtain matching features that lie in the same plane. These matching pairs are important cues to tell the display and the background. We assume that the display is the dominant plane in the scene so that most of the matching pairs are from the plane of the electronic display as shown in Figure 5(a). The use of RANSAC provides robustness to the presence of outliers from the background.

The next step is the initialization of line detection using a kernel-based line detection method [13] as shown in Figure 5(b). The initial detection result is then refined by matching features from the display plane. The goal is to re-

move outlier lines and obtain the candidates lines that correspond the display borders.

Denote M points of RANSAC matching pairs on the first frame as $(x_i^1, y_i^1), i \in \{1, \dots, M\}$. There are N initial lines from Hough line detection phase: $y = \omega_j x + b_j, j \in \{1, \dots, N\}$. If some line is a border, the inliers of M points must be on the same side of this line. In other words, the inliers from $\omega_j x_i^1 + b_j, i \in \{1, \dots, M\}$ should have the same sign. Considering that there are relatively few outliers of matching features, we set a small error tolerance ϵ , typically 0.05. Then the refinement results are those lines with the ratio more than $1 - \epsilon$ of the matching features on one side.

To recognize four borders of the electronic display, we identify horizontal and vertical lines first by their line slopes. Once we have horizontal lines, we can identify the bottom and top lines by the signs of the matched feature points. For example, for the top line, the inliers from $\omega_j x_i^1 + b_j, i \in \{1, \dots, M\}$ are negative while bottom lines have them positive. And we can recognize left and right lines with the similar method. The result is illustrated in Figure 5(c) which shows the candidate lines. There are still outliers from background or the displayed image around borders.

Finally we use an optimization method to get inner borders of the monitor. Specifically we optimize the number-to-area ratio defined as follows. The area is that of the quadrilateral constructed by four types of lines: top, right, bottom and left. The number is the count of matched features within this area. We iterate over all the combinations of four-type lines and select the group with the maximized number-to-area ratio. The final result is shown in Figure 5(d).

Frame-to-Frame Alignment The message embedding described in Section 2.1 has a pointwise division that only works if there is a subpixel alignment between adjacent frames. Consequently, precise image alignment is of great importance. In our work, we use a robust image alignment method with matching features from two neighboring frames, for example $i_o(t)$ and $i_o(t+1)$. The results of aligning $i_o(t+1)$ to $i_o(t)$ are illustrated in Figures 6.

Given two neighboring frames $i_o(t)$ in Figure 6(a) and $i_o(t+1)$ in Figure 6(b), the task is to align $i_o(t+1)$ to $i_o(t)$. We first extract N feature points from $i_o(t+1)$ with locations $(x_i^A, y_i^A), i \in \{1, \dots, N\}$. The Lucas-Kanade tracker algorithm finds their correspondences in $i_o(t)$ with locations $(x_i^B, y_i^B), i \in \{1, \dots, N\}$.

Assume $i_o(t+1)$ transforms into $i_o(t)$ via a perspective warping H and we can represent relationship of feature

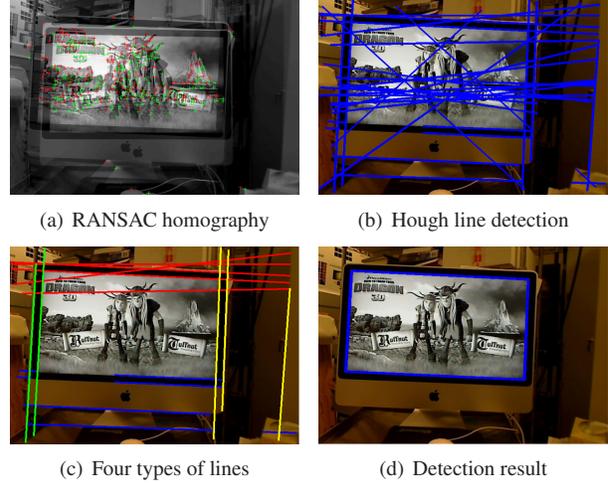


Figure 5. **Automatic display detection results.** RANSAC homography is used to find the matching features from the same plane in 5(a). The initialized line detection is implemented by Hough transformation in 5(b). Line refinement is based on the locations of the matching features relative to the line candidates in 5(c). 5(d) is the display detection result via optimal number-to-area ratio method.

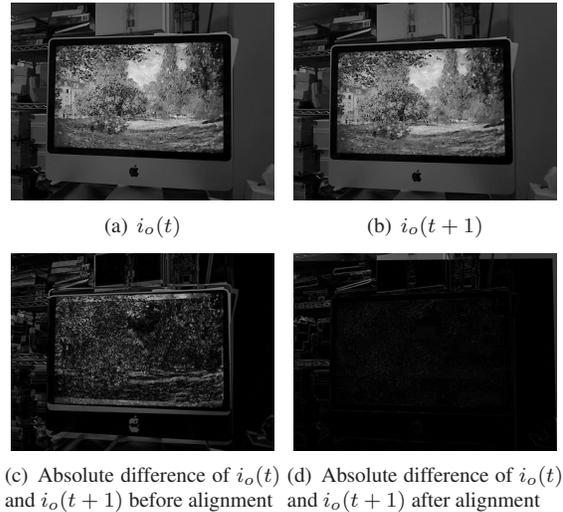


Figure 6. Image alignment of $i_o(t+1)$ to $i_o(t)$

points from these two images with

$$\begin{pmatrix} x_i^B \\ y_i^B \\ 1 \end{pmatrix} = H \begin{pmatrix} x_i^A \\ y_i^A \\ 1 \end{pmatrix}, H = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix}.$$

Reshape H into a vector h where

$$h = (a, b, c, d, e, f, g, h)^T.$$

For the i^{th} feature, we have $R_{\{i,i+1\}} =$

$$\begin{pmatrix} x_i^A & y_i^A & 1 & 0 & 0 & 0 & -x_i^B x_i^A & -x_i^B y_i^A \\ 0 & 0 & 0 & x_i^A & y_i^A & 1 & -y_i^B x_i^A & -y_i^B y_i^A \end{pmatrix},$$

and $c_{\{i,i+1\}} = (x_i^B, y_i^B)^T$. The equation for the i^{th} feature is $R_{\{i,i+1\}}h = c_{\{i,i+1\}}$. Concatenate N points into a $2N \times 8$ matrix named R and a vector c with $2N$ entries. The final equation can be represented as $Rh = c$.

We solve this equation by least square algorithm and the solution is $h = (R^T R)^{-1} R^T c$. Then vector h is reshaped into the homography matrix H . The absolute difference of is shown in Figure 6(c). The absolute difference of $i_o(t)$ and the warped $i_o(t + 1)$ is illustrated in Figure 6(d).

Rectification Rectification is composed of three phases: affine, metric and rotation rectification. Here we use the width-to-height ratio (5 : 3) of the monitor. More details are in [9]. Results are illustrated in Figure 7.

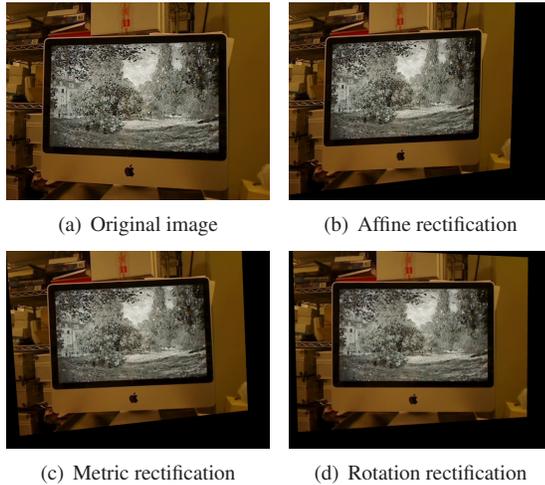


Figure 7. Image rectification

3. Results

We use six display images shown in Figure 9. Each display image is converted to its gray scale and displayed on an iMac with the resolution of 1200×1920 . We play a sequence of 40 frames for each display image: the odd indexed frames are original images and the even indexed ones are embedded with message. The message for each frame is an on/off pattern on a grid that carries the message. The first column of the grid carries a fixed alternating on/off pattern and this fixed pattern is used for training the message bit classifier. In the experiments, we use a PC with Intel CORE i5 and 4G RAM.

We test our system in different environments: indoor and under sunlight. For the sunlight case, specularities is a big



Figure 9. Display images

challenge and RANSAC homography is used to track features from the same plane. Images for the sunlight case is illustrated in Figure 10 and accuracy is listed in Table 2.

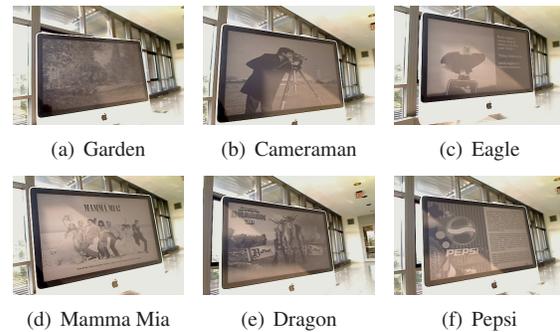


Figure 10. Display images under sunlight

Images	Accuracy(%)	
	indoor	sunlight
Garden	97.6	80.0
Cameraman	88.4	71.3
Eagle	97.1	80.0
Mamma Mia	95.4	86.9
Dragon	96.3	61.9
Pepsi	94.7	77.8
Average	94.8	76.3

Table 2. Table of accuracy for indoor and under sunlight cases

To accelerate speed, the processed image $i_o(t)$ is sub-sampled. The average accuracy of 6 testing sequences is listed in Table 3. The average speed and bitrate are listed in Table 4. As expected, downsampling the processed image reduces the processing time expense. For example, each neighboring pair takes 0.014s with the original size but the time expense reduces to 0.008s for the quarter size. However, the improvement in speed has the cost of accuracy. For example, the display image *Garden* has an accuracy of 97.6% for the full size, but the accuracy reduces to

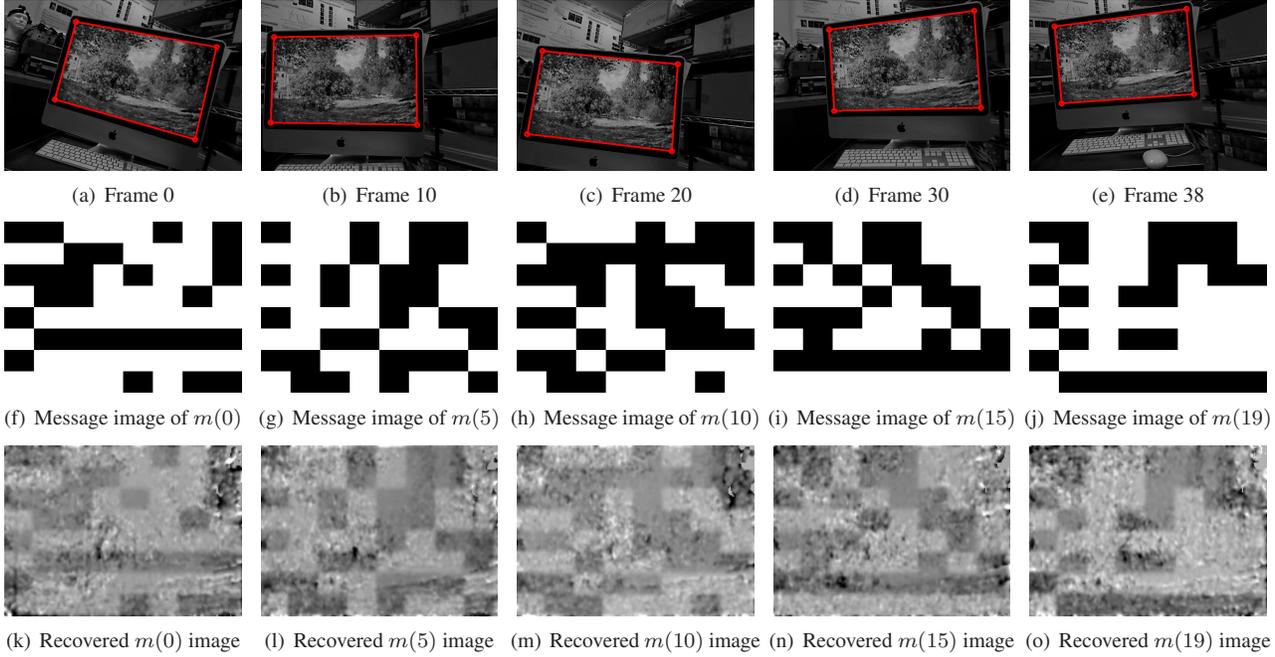


Figure 8. **Extract message images.** Top level: display detection results for the observed images. Middle level: original message images. Bottom level: recovered message images.

78.6% for the quarter size. By observing Table 3, notice the system achieves an average accuracy of 94.6% at the bitrate of 6222.2bps.

Images	Accuracy(%)		
	Full	Half	Quarter
Garden	97.6	96.0	78.6
Cameraman	88.4	87.7	81.2
Eagle	97.1	97.1	76.5
Mamma Mia	95.4	95.8	81.7
Dragon	96.3	96.5	60.7
Pepsi	94.7	93.8	64.3
Average	94.8	94.6	73.8

Table 3. Accuracy for different sizes of processed images

	Full	Half	Quarter
Speed(s/pair)	0.014	0.009	0.008
Bitrate(bits/s)	4000.0	6222.2	7000.0

Table 4. Speed and bitrate for different sizes of processed images

In the visual MIMO system, bit rate can be improved by increasing the number of blocks in the base grid for each message. The results of several block sizes on the message grid are shown in Table 5. Increasing the block size allows more bits to be transmitted, but at the tradeoff of accuracy. Larger block size gives more redundancy because

more pixels transmit the same information. However, as shown in Table 5, correct bit classification of over 90% can be achieved with 16x16 grid (8888.9 bps bitrate) depending on the choice of display image.

Images	Accuracy(%)	
	8 × 8	16 × 16
Garden	97.6	91.3
Cameraman	88.4	72.4
Eagle	97.1	79.6
Mamma Mia	95.4	84.8
Dragon	96.3	85.8
Pepsi	94.7	76.7
Average	94.8	81.8

Table 5. Accuracy for different grid sizes (full size observation images).

	8 × 8	16 × 16
Speed(s/pair)	0.014	0.027
Bitrate(bits/s)	4000.0	8888.9

Table 6. Bitrate for different grid sizes (full size observation images).

4. Acknowledgement

This work is supported by the US National Science Foundation under grant CNS-1065463.

5. Conclusion

In this paper, we present real-time photographic message extraction for visual MIMO optical systems. The method is an interdisciplinary approach that combines methods from computer vision (tracking, object detection, segmentation) with methods from communications (MIMO). While camera-scene communications takes place with bar-codes or QR-codes, this work is unique in that the codes are both dynamic and invisible. The message for transmission is embedded and the electronic display can be simultaneously used for another display purpose. This communication channel can be used with dedicated hardware, in areas with existing electronic signage (e.g. times square, train stations, airports) and in applications with projected imagery (e.g. conference talks, cinema). The initial prototype achieves an accuracy rate at 94.6% at a bitrate of 6222.2 bps. Substantial increase in the bit rate can be realized in future work by improving the accuracy of estimating the projected signal from the received signal and increasing the number of blocks per frame. In the limit a single block can be a single resolvable pixel indicating at 30 frames per second are feasible under conditions of limited spatial blur. Additionally, future developments may also include transmitting multiple bits per block instead of on-off keying.

References

- [1] M. Alghoniemy and A. H. Tewfik. Geometric invariance in image watermarking. *IEEE Transactions on Image Processing*, 13:145 – 153, 2004. 347
- [2] A. Ashok, M. Gruteser, N. Mandayam, T. Kwon, W. Yuan, M. Varga, and K. Dana. Rate adaptation in visual mimo. *Communications Society Conference on Sensor, Mesh, and Ad Hoc Communications and Networks*, 2011. 345, 346
- [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. 345, 346
- [4] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532 – 540, apr 1983. 348
- [5] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 1997. 348
- [6] P. Dong, J. G. Brankov, N. P. Galatsanos, Y. Yang, and F. Davoine. Digital watermarking robust to geometric distortions. *IEEE Transactions on Image Processing*, 14:2140 – 2150, 2005. 347
- [7] M. Farzam and S. S. Shirani. A robust multimedia watermarking technique using zernike transform. in *Proc. IEEE Int. Workshop Multimedia Signal Processing*, 2001. 347
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, pages 125–127. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 348
- [9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, pages 32–37. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 350
- [10] ISO/IEC18004. *Information technology, Automatic identification and data capture techniques, Bar code symbology, QR Code*, 2000. 346
- [11] K. Kamijo, N. Kamijo, and Z. Gang. Invisible barcode with optimized error correction. In *IEEE International Conference on Image Processing (ICIP)*, pages 2036 –2039, oct. 2008. 346
- [12] H. S. Kim and H. K. Lee. Invariant image watermark using zernike moments. *IEEE Transactions on Circuits and Systems for Video Technology*, 13:766 – 775, 2003. 347
- [13] A. F. F. Leandro and M. O. Manuel. Real-time line detection through an improved hough transform voting scheme. *Pattern Recognition*, pages 299–314, 2008. 348
- [14] C.-Y. Lin and S.-F. Chang. Watermarking capacity of digital images based on domain-specific masking effects. In *IEEE Intl. Conf. on Info. Tech: Coding and Computing*, 2001. 347
- [15] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar. Bokode: imperceptible visual tags for camera based interaction from a distance. In *ACM SIGGRAPH*, pages 98:1–98:8, 2009. 346
- [16] S. Nayar. Reflectance ratio: A photometric invariant for object recognition. *IEEE International Conference on Computer Vision*, pages 280–285, 1993. 348
- [17] S. Pereira, S. Voloshynovskiy, and T. Pun. Effective channel coding for dct watermarks. In *ICIP*, pages 671–673, 2000. 347
- [18] 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. 346
- [19] H. Pi, H. Li, and H. Li. A novel fractal image watermarking. *IEEE Transactions on Multimedia*, 8:488 – 499, 2006. 347
- [20] V. Potdar, S. Han, and E. Chang. A survey of digital image watermarking techniques. In *IEEE Intl. Conf. on Industrial Informatics*, 2005. 347
- [21] C. W. Tang and H. M. Hang. A feature-based robust digital image watermarking scheme. *IEEE Transactions on Signal Processing*, 51:950 – 959, 2003. 347
- [22] M. Varga, A. Ashok, M. Gruteser, N. Mandayam, W. Yuan, and K. Dana. Demo: visual mimo based LED - camera communication applied to automobile safety. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 383–384, 2011. 346
- [23] S. Xiang, H. J. Kim, and J. Huang. Invariant image watermarking based on statistical features in the low-frequency domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 18:777 – 790, 2008. 347
- [24] W. Yuan, K. Dana, M. Varga, A. Ashok, M. Gruteser, and N. Mandayam;. Computer vision methods for visual mimo optical system. *IEEE Computer Vision and Pattern Recognition Workshop(ProcAMS)*, 2011. 345, 346