# Reconstruction of Gesture Images by Using Banner as Illumination of Single-Pixel Imaging

Naoya Mukojima[1], Hiroki Takatsuka[1], Masaki Yasugi[1], Shiro Suyama[1], Hirotsugu Yamamoto[1]

hirotsugu@yamamotolab.science
[1]Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya, Tochigi, 321-0904, Japan
Keywords: single-pixel imaging, apparent image, encoded image

## ABSTRACT

*This paper proposes a method to reconstruct a gesture image by single-pixel imaging (SPI) using a banner displayed on an electronic bulletin board as illumination. We have succeeded in reconstructing the gesture by SPI using 1500 randomly encoded sub-frames of the banner.*

## 1 Introduction

Single-pixel imaging is a method which can reconstruct images by using a point photodetector and variable illuminations [1], and is suitable gesture recognition by reconstructing shadow pictures in considering of privacy [2]. For an application in public spaces, we have proposed single-pixel imaging using banners on electronic billboards as illumination [3]. When an arbitrary region is set, a different display will appear within that region each time the banner display flows. We investigated the use of this method as illumination in single-pixel imaging. But in image reconstruction using simple banners, there is a problem that the appearance of the text on the banner is superimposed on the reconstructed image. To solve the problem, we propose an image reconstruction method that uses a random pattern latent in the apparent image on a high-frame-rate LED display for banner illumination [4].

The purpose of this paper is to clarify the possibility of reconstructing a single-pixel imaging gesture image using a binary image banner as the apparent image and a randomly encoded sub-frame of image as the illumination.

## 2 Principle

### 2.1 Single-Pixel Imaging

The diagram of single-pixel imaging is shown in Fig. 1. The target object of this study is a subject's hand. The illumination is modulated an arbitrary number of times by a randomly generated mask pattern. Summation of the transmitted light intensity is detected by a point photodetector. By calculating the correlation between the illumination intensity and the detected intensity, it is possible to obtain a reconstructed image as a result of floating-point arithmetic [5]. The detected intensity at by $k$-th illumination, denoted by $\mathrm{B}_k$ and obtained by a single-pixel detector, can be expressed as:

$$\mathrm{B}_k = \iint_S \mathrm{I}_k(x,y)\mathrm{T}(x,y)\,\mathrm{d}x\,\mathrm{d}y\,. \qquad (1)$$

where $\mathrm{I}_k(x,y)$ represents the light intensity of $k$-th random mask pattern at the position $(x,y)$ ; $\mathrm{T}(x,y)$ is the transmittance of the target object on $(x,y)$; $\mathrm{I}_k(x,y)$ is given in advance for $k = 1,2,\ldots n$; $n$ is the total number of illuminations; the integration of $\mathrm{B}_k$ is carried out over the area of the object $S$. In order to reconstruct the object image, we use the correlation function G. G approaches $\mathrm{T}(x,y)$ with increasing the number of illumination $n$ [6]. G can be expressed as:

$$\mathrm{G}(x,y) = \langle \Delta\mathrm{B}_k \Delta\mathrm{I}_k(x,y)\rangle. \qquad (2)$$

$\langle\ \rangle$ represents the ensemble, which is the average of $n$ consecutive measurements. $\Delta\mathrm{I}_k(x,y)$ is the deviation between the light intensity of the $k$-th random mask pattern and the ensemble. $\Delta\mathrm{B}_k$ is the deviation between the light intensity of the $k$-th measurement and the ensemble. If $N$ consecutive measurements are denoted by $\mathrm{F}_i$, the ensemble can be expressed as:

$$\mathrm{F}_i = \frac{1}{N}\sum_{i=1}^{N} \mathrm{F}_i \qquad (3)$$

The reconstruction results of the gesture image at each measurement count is shown in Fig. 2.

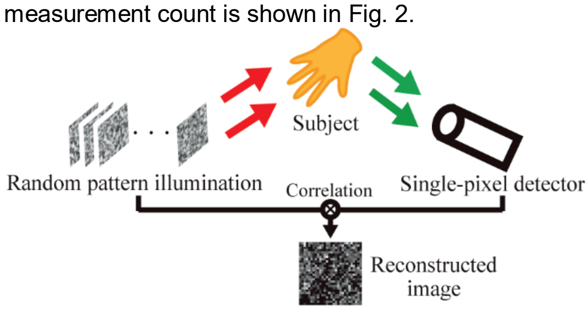


**Fig. 1 Diagram of single-pixel imaging to detect a silhouette image of a subject's hand gesture**

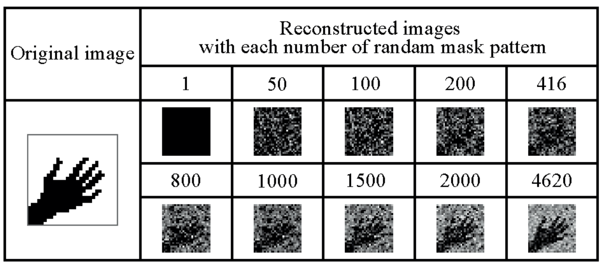


**Fig. 2 Example of reconstructed images with each number of random mask pattern**

### 2.2 Single-Pixel Imaging Using Banner Image

Banner images were used instead of random mask patterns in SPI. Fig. 3 shows how to extract a banner image as a mask image. The banner image is an image of the same height as the original image (32 × 32 pixels). The first mask image is extracted from the banner image horizontally from 0-th to 31-st. In this case, the height is the same, so the mask image will be 32 × 32 pixels. For the second and subsequent mask images, the area of interest is shifted by 1 pixel toward the edge. In this method, the total number of mask patterns is determined by the width of the banner image.

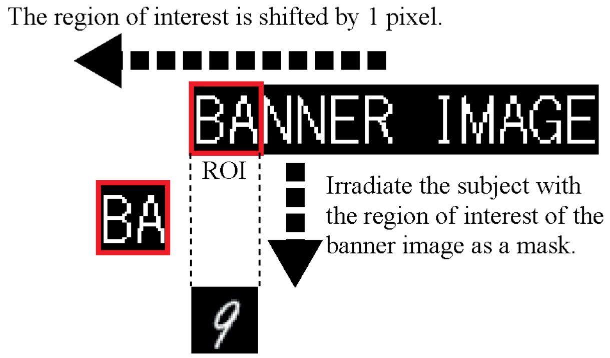


**Fig. 3 Extraction of a banner image as a mask image**

### 2.3 Banner Image Encoding

In addition to the method described in section 2.2, the banner image used as the mask image is a randomly encoded sub-frames of image. For the encoding of banner images, the pixel values of the banner images used in this paper are set to 70 and 255 before encoding. This pixel value is based on the assumption of the actual display for an electronic bulletin board and the brightness of the display board. The number of images to be encoded is 15. The number of images to be encoded is 15, and the pixel composition of each encoded image is shown in Table 1. The order in which the pixel values are displayed is random.

**Table 1 Values of each pixel by encoding**

| Original pixel value | Number of pixel value 0 | Number of pixel value 10 | Number of pixel value 185 | Total encoded images |
|---|---|---|---|---|
| 70 | 8 | 7 | 0 | 15 |
| 255 | 7 | 7 | 1 | 15 |

The 15 coded images with the pixel configurations shown in Table 1 are displayed at high speed on a high-frame-rate LED display, resulting in a value equal to the original pixel value.

The single-pixel imaging with encoded images was simulated assuming the use of a high-frame-rate LED display (screen switching time: 1 ms), so the flow time of the banner image was considered. Accordingly, the banner flow time was set to 5 characters / sec for readability [7]. The banner image used in this paper has a size of 32 × 416 pixels and consists of 26 characters, with the width of each character being 16 pixels. When the banner moves at 5 characters / sec, 80 pixels / sec will flow. In other words, it takes 12.25 ms to moves 1 pixel. Since the screen switching time of a high-frame-rate LED display is 1 ms, approximately 12 encoded images can be displayed as a mask image moves 1 pixel. In this paper, 12 images / ms were used as a simulation.

However, 15 encoded images are required to display the banner image as an apparent image. For Example, the remaining 3 encoded images that could not be displayed during the 1 ms period between 0 ms and 1 ms are displayed as mask images after the ROI moves over the 1-pixel banner image after 1 ms. In other words, only 12 banners can be displayed in 1 ms, so the banner cannot be a real apparent image within the region of interest alone. However, the subject can actually see the banner as an apparent image because the entire banner is displayed to the subject and the region of interest is see only within the measurement system. The encoded images used at each time are shown in Fig. 4. Compared to the conventional image reconstruction using banner images, this method allows 12 sub-frames of image to be displayed before moving 1 pixel, increasing the number of measurements when using the same size banner image by a factor of 12 in the case of this study.

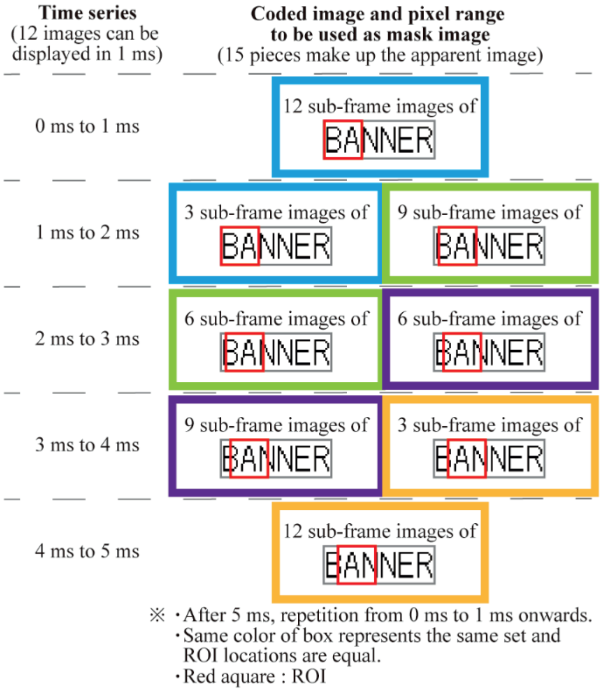


**Fig. 4 Details of encoded images used for each time series**

## 3 Reconstruction of Result with Conventional Methods

### 3.1 Experiment

The banner image size was 32 × 416 (8 bit), and the original image for reconstruction were MNIST [8], a dataset of handwritten numbers from 0 to 9. 10 images were randomly selected from each number in MNIST. The original image size was 32 × 32 (8 bit). Using the banner image in Fig. 5, image reconstruction was performed using the methods in section 2.2. In order to compare the results with those obtained using a normal banner, we reconstructed the image using the banner image in (c) of Fig. 5 and use the method in section 2.3.


**Fig. 5 Banner images. (a) random mask pattern, (b) a string of uppercase only "A", (c) a string of uppercase letters from A to Z, (d) a string of lowercase letters**

### 3.2 Results

The reconstruction results are shown in Fig. 6. Also the reconstruction results using the sub-frame image encoded from the banner image in (c) of Fig. 5 as mask images are shown in Fig. 6.

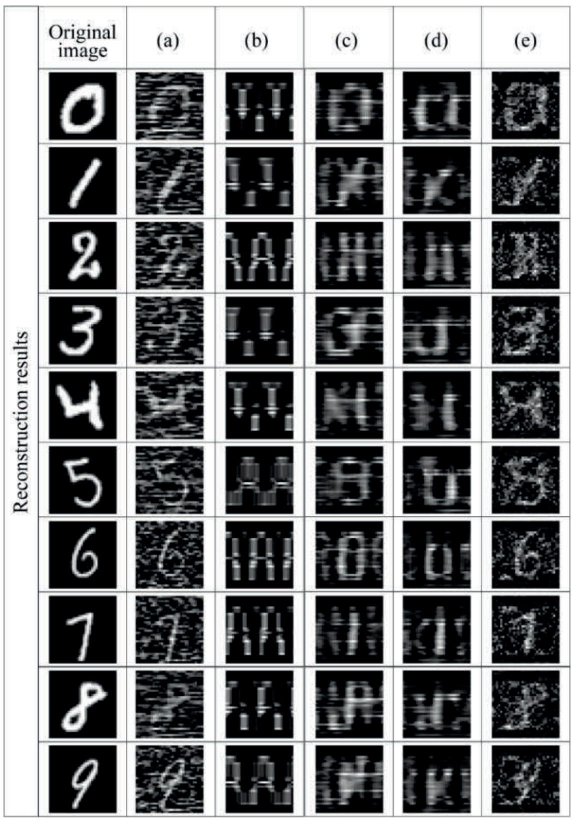


**Fig. 6 Reconstruction of result using uncorded banner images. (a) random mask pattern, (b) a string of uppercase only "A", (c) a string of uppercase letters from A to Z, (d) a string of lowercase letters, (e) randomly encoded sub-frame images in (c)**

The results of (a) shows that the image can be reconstructed when a random pattern is used. However, for results (b) through (d), the apparent letters of each banner were superimposed on the reconstructed image, and the original image of each number could not be reconstructed. The problem of super imposition of the appearance of letters was particularly noticeable in (b). The result of (e), although not as good as (a), is better than that of the method using the usual banner image, and the number of reconstructed results that can be confirmed for each number increases.

## 4 Reconstruction of Results Using the Banner Images as a Randomly Encoded Sub-Frames of Image

### 4.1 Experiment

To solve the problem where the apparent text is superimposed on the reconstructed image, we reconstructed the image using the randomly encoded sub-frames image of the banner image shown in 2.3 as a mask image. The banner images used are (b) through (d) in Fig. 5, and Fig.7 shows an example of sub-frames of image of (b) that was actually randomly encoded, 32 × 416 pixels, 8bits. For the purpose of this paper, the original image is a gesture image in considering of privacy, 32 × 32 pixels, 8bits. The number of measurements was 1, 50, 100, 200, 416, 800, 1000, 1500, 2000, and 4620. The 4620 times is calculated by multiplying 385, the maximum number of measurements for image reconstruction using conventional uncoded banners, by 12, the number of images that can be displayed in 1 ms by using encoded random sub-frames of image.

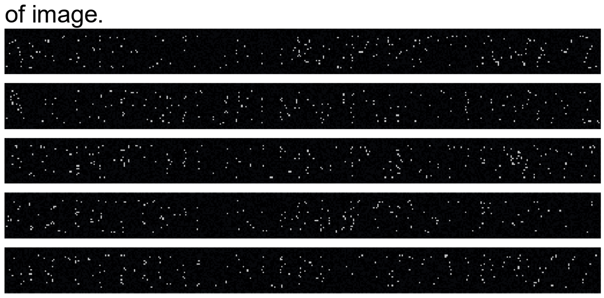

**Fig. 7 Example of sub-frames of image**

### 4.2 Results

The reconstruction results are shown in Fig. 8. It was show that the original image was reconstructed after 1500 measurements when a randomly encoded sub-frames of image used as mask images and when the character strings composing the banner were different. The result of (b) shows that the apparent text is superimposed on the original image, and that the overall gesture is elongated compared to the gesture in the original image. In (c) to (d), we found that the apparent superimposition of letters on the reconstructed image, which was identified in Section 3.2, was reduced by comparing the results.

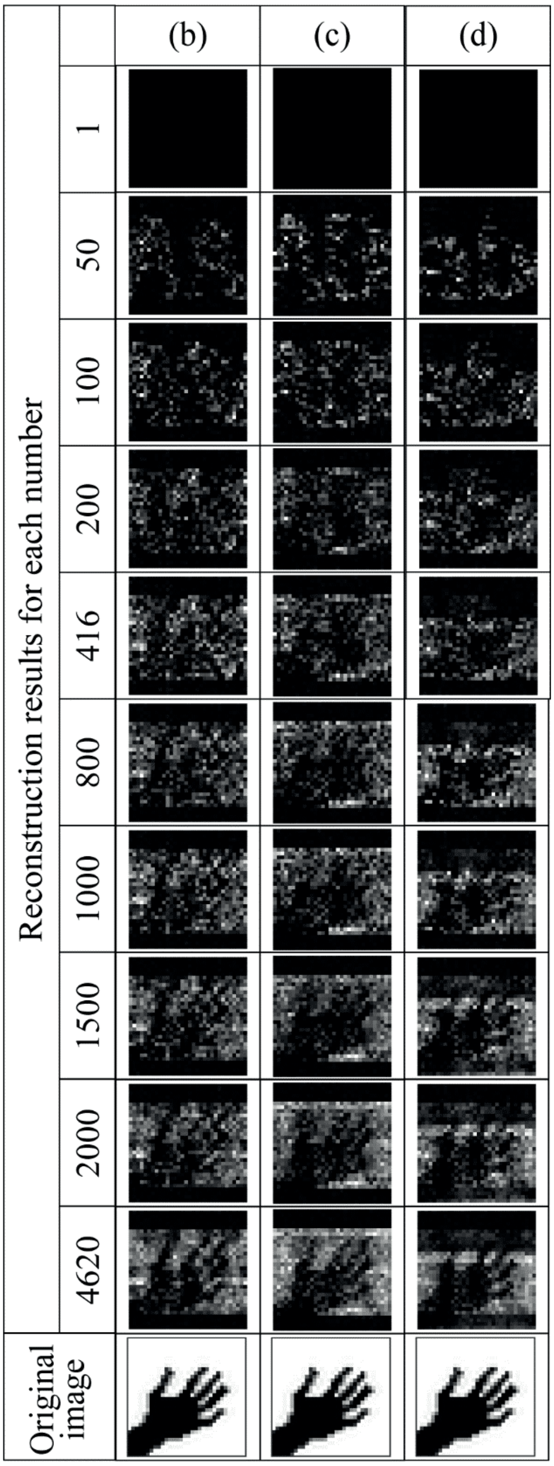


**Fig. 8 Reconstruction of results using the banner images as a randomly encoded sub-frames of image**

## 5 Discussion

The result in (b) of section 4.2 shows that the gesture is stretched. This is due to the apparent superimposed of letters. The problem of the apparent superimposition of letters is easily influenced by the pixel values of the letters, especially when the number of measurement is small. In addition, the effect of the letters becomes larger when the width of the letters is larger, so this problem can be solved by changing the font and the width of the letters. Also in this paper, we found that more than 1500 measurements were required, but a faster number of reconstructions is desirable for gesture recognition. Therefore, our proposed method [9], which restore a reconstructed image with a small number of measurements by deep-learning. This method can be used to reconstruct a gesture image with fewer than 1500 measurements.

## 6 Conclusion

In this paper, we investigated the possibility of image reconstruction using a method of latent random patterns in apparent images on high-frame-rate LED displays for displaying banner. The original image is reconstructed after 1500 measurements when the character strings that compose the banner are different.

This show the possibility of reconstructing a gesture image by single-pixel imaging using the proposed banners as an apparent image and a randomly encoded sub-frames of image as illumination.

## References

[1] J.H. Shapiro, “Computational ghost imaging,” Phys. Rev. A - At. Mol. Opt. Phys., Vol. 78, 061802 (2008).
[2] N. Mukojima, M. Yasugi, Y. Mizutani, T. Yasui, H. Yamamoto, “Deep-Learning-Assisted Single-Pixel-Imaging for Gesture Recognition in Considering for Privacy,” IEICE. Transaction on Electronics., Vol. E105-C, No. 2, pp. 79–85 (2022).
[3] N. Mukojima, M. Yasugi, S. Suyama, H. Yamamoto, “The Possibility of Using Banner Images as the Mask Pattern of Single-Pixel Imaging,” Proc. IP’22, IPp-09 (2022).
[4] M. Takahashi, H. Yamamoto, “Encryption by spatiotemporal scrambling on a high-frame-rate display,” The 63rd JSAP Spring Meeting, 21a-S224-5 (2016) [in Japanese].
[5] K. Shibuya, K. Nakae, Y. Mizutani, and T. Iwata, “Comparison of reconstructed images between ghost imaging and Hadamard transform imaging,” Opt. Rev., Vol. 22, pp. 897–902 (2015).
[6] Y. Bromberg, O. Katz, and Y. Silberberg, “Ghost imaging with a single detector,” Phys. Rev. A, Vol. 79, 53840 (2009).
[7] K. Kanaoka, R. Nakayama, K. Notomi, K. Saito, “Fundamental examination of optimal ticker scroll speed for reading newsflash,” J. FIT’09, J-051 (2009) [in Japanese].
[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, Vol. 86, No. 11, pp. 2278-2324 (1998).
[9] N. Mukojima, M. Yasugi, Y. Mizutani, T. Yasui, H. Yamamoto, “Reducing the Number of mask Patterns in Single-Pixel Imaging Using Deep Learning,” The 68th JSAP Spring Meeting, 16a-P07-3 (2021) [in Japanese].