Title

Integrated Computational Modeling for Accelerated Single-Pixel Imaging

Industrial Partner:

Information Technology R&D Center, Mitsubishi Electric Corp.

Mitsubishi Electric is a world leader in manufacturing and sales of electrical and electronic products and systems used in a broad range of fields and applications. As a global leader among green companies, our technologies are being applied to contribute to and support society and daily life around the world. The Information Technology R&D Center is actively creating new businesses through basic research and development in the fields of information technology, media intelligence, electro-optics microwaves, and communication technologies. We are also seeking technologies that reinforce our position on the leading edge of progress, with work to renew existing businesses through the fruits of our R&D in the field of IT.

Industry Mentors

Keito Arifuku, Ph.D., Head Researcher, Computational Modeling and Analysis Group, Advanced Basic Research Dept.

Ryuhei Takahashi, Ph.D., Senior Manager, Computational Modeling and Analysis Group, Advanced Basic Research Dept.

Background

Single-pixel imaging (SPI) is a technique that reconstructs images from measurements using modulated illumination and a single-pixel detector, enabling high-sensitivity and low-cost imaging even beyond the visible spectrum or under extremely low-light conditions. However, it generally requires sequential projection of a large number of modulation patterns, making acquisition speed a bottleneck. Introducing unsupervised deep learning into SPI reconstruction can reduce the number of required samples and accelerate recovery, but the ill-conditioned nature of the inverse problem derived from the measurement equation, coupled with the absence of direct supervisory signals in unsupervised learning, may degrade reconstruction accuracy. In this study, we investigate an optimization approach that integrates mathematical regularization and physical models into deep learning to reduce performance. The proposed method will be evaluated using reconstruction accuracy, robustness to noise variations, and acquisition speed as key metrics.

Principle

SPI is an imaging technique that reconstructs a 2D image of an object using only one light-sensitive detector (a single pixel) instead of a conventional camera with many pixels. As shown in Fig. 1, digital micromirror device (DMD) in SPI can be realized by two schemes [1], that is, the structured illumination (front modulation) and the structured detection (back modulation). In the structured illumination scheme, the DMD is placed before the object to project structured patterns, and the back-scattered light is collected by a single-pixel detector. The field of view is determined by the illumination area, while the shading profile depends on the detector position, consistent with Helmholtz reciprocity [2] (or dual photography [3]). In contrast, in the structured detection scheme, the DMD is placed after the object to modulate the detected light field. Here, the field of view is defined by the detection components, and the shading profile is determined by the illumination source. Differences in the point spread function between the two schemes are discussed in [4].

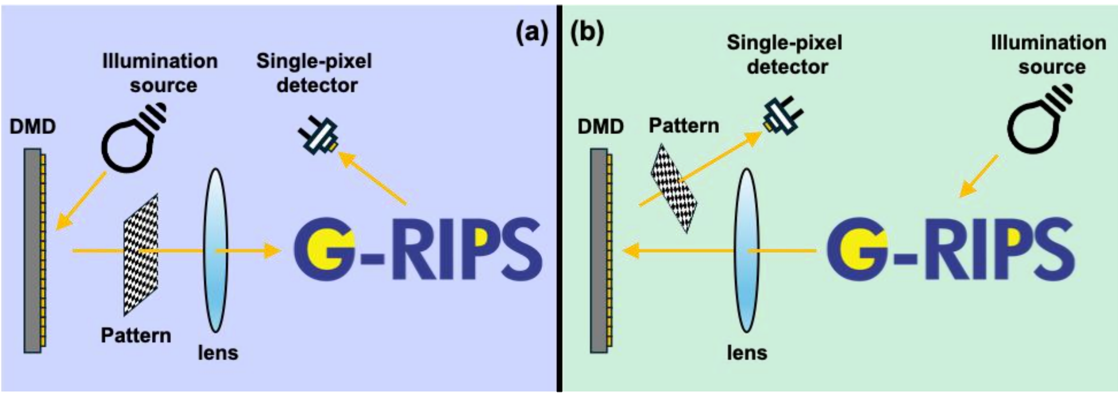

Fig.1 Two different spatial light modulation schemes in SPI. (a) Structured illumination scheme, where the DMD is placed before the object and used to modulate the illumination light field; (b) structured detection scheme, where the DMD is placed after the object and used to modulate the detection light field.

In the physical model, let $f(x, y)$ denote the luminance distribution of the target object and $P_i(x, y)$ denotes the i-th illumination pattern generated by a spatial light modulator (such as a DMD). The light intensity signal y_i captured by the detector can be expressed as:

$$y_i = \iint f(x,y)P_i(x,y)dxdy + n_i \,, \quad (1)$$

where n_i denotes additive observation noise. By discretizing the image into an N-dimensional vector $\mathbf{x}$ and the patterns into an $M\times N$ observation matrix A, the fundamental equation is described as $\mathbf{y} = A\mathbf{x} + \mathbf{n}$. The differences among the SPI methods mainly fall into three categories: 1) the type of structured patterns used in spatial light modulation, 2) the sampling strategy (specifically, the ordering of structured patterns), and 3) the image reconstruction algorithm.

Statistical Correlation Methods

SPI based on spatial light modulation originates from ghost imaging (GI), first demonstrated in 1995 using entangled photons [5]. Subsequent studies showed that GI can be realized with classical light sources and is fundamentally governed by classical coherent propagation, enabling imaging with a single-pixel detector that measures only intensity information. So, GI reconstructs images based on the correlation between the intensity distribution of illumination patterns and the total light intensity detected. The advanced GI with the use of a spatial light modulator to generate programmable speckle patterns is proposed and called computational ghost imaging (CGI) [6]. CGI utilizes known random patterns P_i generated by a computer. As Eq. (2) shows, CGI uses a correlation-based algorithm for image reconstruction and the algorithm is iterative.

$$\hat{\mathbf{x}} = \langle (y_i - \langle y \rangle) P_i \rangle \, , \tag{2}$$

where image $\hat{\mathbf{x}}$ is calculated through the ensemble average of the observation y_i and the pattern fluctuations. As is based on a statistic model, CGI typically requires far more single-pixel measurements than image pixels (that is, $n \gg M \times N$) to reproduce a high-quality image. For example, Sun et al. reconstructed a 128×96-pixel image with 10^6 single-pixel measurements [7].

Differential ghost imaging (DGI) was proposed to suppress the effects of light source fluctuations and background noise [8]. By introducing the total intensity of each pattern, $S_i = \iint P_i(x,y)dxdy$, the following operation is performed:

$$\hat{\mathbf{x}}_{\mathrm{DGI}} = \langle \left(y_i - \frac{\langle y \rangle}{\langle S \rangle} S_i \right) P_i \rangle \, . \tag{3}$$

This differential processing improves sensitivity to signal components, enabling a significantly higher signal-to-noise ratio (SNR) compared to conventional GI. Same with CGI, a major limitation of differential ghost imaging is that reliable estimation of the mean values requires a large number of measurements, leading to increased acquisition time, while its performance may degrade under low sampling ratios or in the presence of strong nonlinear responses and spatially varying noise.

Optimization via Compressive Sensing

Compressive sensing ghost imaging (CSGI) is an innovative imaging technique that merges the optical correlation principles of GI with the mathematical framework of compressive sensing (CS) [9]. Traditional ghost imaging requires a vast number of measurements often far exceeding the total pixel count to reconstruct a clear image, leading to long acquisition times. CSGI overcomes this limitation by treating image reconstruction as an optimization problem that leverages the inherent "sparsity" of natural scenes. Following the fundamental SPI model described in the previous section, the power of CSGI lies in its ability to recover the target image vector **x** even when the system is highly underdetermined ($M \times N \gg n$). Although a raw image **x** is typically dense, it becomes “sparse” when transformed into an appropriate basis Ψ, such as wavelet or discrete cosine transform [10]. Defining the relationship as $\mathbf{x} = \Psi\alpha$, where α is a sparse coefficient vector, the missing information is reconstructed by solving the following L_1-norm minimization problem:

$$\min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \|\mathbf{y} - A\Psi\alpha\|_2^2 < \epsilon \, . \tag{4}$$

By solving this convex optimization problem, high-quality images can be reconstructed from sub-Nyquist sampling rates often using only 20% to 30% of the data required by traditional methods [11].

Orthogonal Basis Sampling Methods

Instead of random patterns, these methods use mathematically orthogonal bases as patterns to perform deterministic and efficient reconstruction. The Hadamard basis consists of binary matrices with elements {1, -1}, which align well with the binary mirror control of a DMD. After projecting all patterns, applying a fast Hadamard transform (FHT) allows for image reconstruction with minimal computational load as $\mathbf{x} = H^{-1}\mathbf{y}$. The details are discussed in [12]. Hadamard single-pixel imaging (HIS) provides high SNR and computationally efficient reconstruction through the Fast Hadamard Transform, enabling high-quality image recovery at sub-Nyquist sampling rates (20% to 30%) when combined with CS. However, it is constrained by a resolution tied to powers of two and remains highly sensitive to optical alignment and motion blur.

Fourier single-pixel imaging (FSI) directly samples the spatial frequency domain (Fourier domain) of the object [12]. Using sinusoidal patterns, the complex Fourier coefficient $C(f_x, f_y)$ is calculated, typically via the four-step phase-shifting method:

$$C\left(f_x, f_y\right) = \left(y_0 - y_\pi\right) + j\left(y_{\pi/2} - y_{3\pi/2}\right) \ . \tag{5}$$

The image is obtained through an Inverse Fast Fourier Transform (IFFT). By prioritizing low-frequency measurements, FSI achieves extremely high sampling efficiency. FSI offers flexible resolution and superior efficiency for natural images by sampling sparse coefficients in the frequency domain, but it requires complex sinusoidal modulation and multiple phase-shifting steps, making it generally slower and more difficult to implement on binary DMDs compared to the high-speed, binary patterns of HSI. A major challenge in FSI is that the DMD is a binary device (0/1) and cannot directly represent grayscale sinusoidal waves. Binary Fourier single-pixel imaging (BFSI) overcomes this by using dithering algorithms to convert sinusoids into pseudo-binary halftone patterns. Physically, BFSI intentionally utilizes the defocus (blur) of the optical projection system to act as a low-pass filter that removes high-frequency binary noise.

$$y = \iint f(x, y)[B(x, y) * PSF(x, y)]dxdy \ , \tag{5}$$

where $B(x, y)$ is the binary pattern and PSF is the point spread function. This method enables Fourier sampling at the maximum refresh rate of the DMD, significantly contributing to the high-speed operation of SPI [13]. BFSI simplifies standard FSI by using binary dithering or pulse-width modulation to approximate grayscale sinusoidal patterns on high-speed DMDs. While this significantly increases measurement speed and maintains FSI's flexible resolution, it introduces harmonic artifacts and quantization noise that can degrade image quality compared to the pure sinusoidal modulation of conventional FSI.

Expectation

Deep learning (DL) provides a powerful framework for addressing the ill-posed inverse problems in SPI by automatically learning hierarchical features from measurements. Employing architectures such as convolutional neural networks (CNNs) and generative adversarial networks (GANs) within data-driven, model-driven, or hybrid paradigms has substantially alleviated traditional bottlenecks in reconstruction speed and image quality and enabled capabilities such as super-resolution, imaging through scattering media, and image-free sensing [14, 15]. Nevertheless, DL-based SPI still faces critical challenges, including a heavy dependence on large, high-quality training datasets, limited interpretability of learned models, and a fundamental trade-off between real-time processing speed and high-fidelity reconstruction. In addition, conventional SPI techniques suffer from a physical dilemma: achieving high spatial resolution typically requires an extremely large number of sampling patterns (i.e., long effective exposure), and the attainable resolution is ultimately bounded by the diffraction limit of the optical system. To address the combined challenges of extensive sampling requirements, diffraction-limited resolution, and dataset insufficiency or bias, this study adopts a hybrid approach that combines the following three strategies:

1) Integration of a Physical Model into a Deep Neural Network (DNN)

The physical model of SPI image formation is incorporated into a deep neural network, imposing constraints so that the network output is consistent with the physically measured data (bucket signals), rather than relying solely on unconstrained inference.

2) Unsupervised Learning without Pre-training

Since optical sensing inherently involves detecting unknown targets, pre-training is not always suitable for real-time measurements. In this work, we adopt an unsupervised, non-pretrained learning framework that leverages prior information, thereby avoiding the bias caused by dependence on a specific training dataset and enabling flexible adaptation to previously unseen objects.

3) Iterative Optimization Process

A coarse initial image is first generated using a conventional SPI method and fed into the DNN. The network output is treated as the current estimated image, from which the physical model computes the corresponding measurement values. The network weights are iteratively updated to minimize the error (loss function) between the estimated and actual measurements, and mathematical regularization (e.g., Total Variation) is optionally applied to suppress noise and enhance image quality.

The effectiveness of the proposed method will be evaluated using reconstruction accuracy, robustness to noise variations, and acquisition speed as the primary performance metrics. To carry out the hybrid approach and its evaluation, candidates should possess basic knowledge of, or a strong interest in, AI concepts, as well as practical programming skills in Python, including the ability to implement algorithms and utilize scientific libraries.

References

[1] M. P. Edgar, et al., “Principles and prospects for single-pixel imaging,” Nat. Photonics 13, 13-20 (2019).

[2] B. Max, and W. Emil, “Principles of Optics: electromagnetic Theory of Propagation,” Interference and diffraction of light (7th ed.), Cambridge University Press (1988).

[3] P. Sen, et al., “Dual photography,” ACM Trans. Graph. 24, 745-755 (2005).

[4] D. Shi, et al., “Enhancing resolution of single-pixel imaging system,” Opt. Rev. 22, 802-808 (2015).

[5] T. B. Pittman, et al., “Optical imaging by means of two-photon quantum entanglement,” Phys. Rev. A 52, R3429-R3432 (1995).

[6] J. H. Shapiro, and B. I Erkmen, “Ghost imaging: from quantum to classical to computational,” Adv. Opt. Photonics 2, 405-450 (2010).

[7] B. Sun, et al., “3D computational imaging with single-pixel detectors,” Science, 340, 844-847 (2013).

[8] F. Ferri, “Differential Ghost Imaging,” Phys. Rev. Lett. 105, 219902 (2010).

[9] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory 52, 1289-1306 (2006).

[10] D. L. Donoho, and J. Tanner, “Counting faces of randomly projected polytopes when the projection radically lowers dimension,” J. Am. Math. Soc. 22, 1-53 (2009).

[11] A. Kallepalli, et al., “Compressed sensing in the far-field of the spatial light modulator in high noise conditions,” Sci. Rep. 11, 17460 (2021).

[12] Z. Zhang, et al., “Hadamard single-pixel imaging versus Fourier single-pixel imaging,” Opt. Express 25, 19619-19639 (2017).

[13] Z. Zhang, et al., “Fast Fourier single-pixel imaging via binary illumination,” Sci. Rep. 7, 1-9 (2017).

[14] K. Song et al., “Single-pixel imaging based on deep learning,” arXiv, 2310.16869 (2025).

[15] K. Song et al., “Advances and challenges of single-pixel imaging based on deep learning,” Laser Photonics Rev. 19, 2401397 (2025).