

Patch-primitive driven compressive ghost imaging

Xuemei Hu,^{1,2} Jinli Suo,^{1,2} Tao Yue,^{1,2} Liheng Bian,^{1,2} and Qionghai Dai,^{1,2,*}

¹*Department of Automation, Tsinghua University, Beijing 100084, China*

²*Beijing Key Laboratory of Multi-dimension and Multi-scale Computational Photography (MMCP), Tsinghua University, Beijing 10084, China*

[*qhdai@tsinghua.edu.cn](mailto:qhdai@tsinghua.edu.cn)

Abstract: Ghost imaging has rapidly developed for about two decades and attracted wide attention from different research fields. However, the practical applications of ghost imaging are still largely limited, by its low reconstruction quality and large required measurements. Inspired by the fact that the natural image patches usually exhibit simple structures, and these structures share common primitives, we propose a patch-primitive driven reconstruction approach to raise the quality of ghost imaging. Specifically, we resort to a statistical learning strategy by representing each image patch with sparse coefficients upon an over-complete dictionary. The dictionary is composed of various primitives learned from a large number of image patches from a natural image database. By introducing a linear mapping between non-overlapping image patches and the whole image, we incorporate the above local prior into the convex optimization framework of compressive ghost imaging. Experiments demonstrate that our method could obtain better reconstruction from the same amount of measurements, and thus reduce the number of requisite measurements for achieving satisfying imaging quality.

© 2015 Optical Society of America

OCIS codes: (110.1650) Coherence imaging; (110.1758) Computational imaging; (150.1135) Algorithm.

References and links

1. T. Pittman, Y. Shih, D. Strekalov, and A. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Phys. Rev. A* **52**, R3429 (1995).
2. R. S. Bennink, S. J. Bentley, and R. W. Boyd, "Two-photon coincidence imaging with a classical source," *Phys. Rev. Lett.* **89**, 113601 (2002).
3. A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato, "Ghost imaging with thermal light: comparing entanglement and classical correlation," *Phys. Rev. Lett.* **93**, 093602 (2004).
4. A. Valencia, G. Scarcelli, M. D'Angelo, and Y. Shih, "Two-photon imaging with thermal light," *Phys. Rev. Lett.* **94**, 063601 (2005).
5. J. H. Shapiro, "Computational ghost imaging," *Phys. Rev. A* **78**, 061802 (2008).
6. Y. Bromberg, O. Katz, and Y. Silberberg, "Ghost imaging with a single detector," *Phys. Rev. A* **79**, 053840 (2009).
7. B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. Padgett, "3D computational imaging with single-pixel detectors," *Science* **340**, 844–847 (2013).
8. N. Tian, Q. Guo, A. Wang, D. Xu, and L. Fu, "Fluorescence ghost imaging with pseudothermal light," *Opt. Lett.* **36**, 3302–3304 (2011).
9. P. Clemente, V. Durán, E. Tajahuerce, J. Lancis, "Optical encryption based on computational ghost imaging," *Opt. Lett.* **35**, 2391–2393 (2010).

10. W. Chen and X. Chen, "Ghost imaging for three-dimensional optical security," *Appl. Phys. Lett.* **103**, 221106 (2013).
11. C. Zhao, W. Gong, M. Chen, E. Li, H. Wang, W. Xu, and S. Han, "Ghost imaging lidar via sparsity constraints," *Appl. Phys. Lett.* **101**, 141123 (2012).
12. J. Cheng, "Ghost imaging through turbulent atmosphere," *Opt. Express* **17**, 7916–7921 (2009).
13. P. Zhang, W. Gong, X. Shen, and S. Han, "Correlated imaging through atmospheric turbulence," *Phys. Rev. A* **82**, 033817 (2010).
14. O. S. Magaña-Loaiza, G. A. Howland, M. Malik, J. C. Howell, and R. W. Boyd, "Compressive object tracking using entangled photons," *Appl. Phys. Lett.* **102**, 231104 (2013).
15. E. Li, Z. Bo, M. Chen, W. Gong, and S. Han, "Ghost imaging of a moving target with an unknown constant speed," *Appl. Phys. Lett.* **104**, 251120 (2014).
16. F. Ferri, D. Magatti, L. Lugiato, and A. Gatti, "Differential ghost imaging," *Phys. Rev. Lett.* **104**, 253603 (2010).
17. B. Sun, S. S. Welsh, M. P. Edgar, J. H. Shapiro, and M. J. Padgett, "Normalized ghost imaging," *Opt. Express* **20**, 16892–16901 (2012).
18. W. Wang, Y. P. Wang, J. Li, X. Yang, and Y. Wu, "Iterative ghost imaging," *Opt. Lett.* **39**, 5150–5153 (2014).
19. K. W. C. Chan, M. N. O'Sullivan, and R. W. Boyd, "High-order thermal ghost imaging," *Opt. Lett.* **34**, 3343–3345 (2009).
20. W. Gong and S. Han, "Phase-retrieval ghost imaging of complex-valued objects," *Phys. Rev. A* **82**, 023828 (2010).
21. M. Mirhosseini, O. S. Magaña-Loaiza, S. M. H. Rafsanjani, and R. W. Boyd, "Compressive direct measurement of the quantum wave function," *Phys. Rev. Lett.* **113**, 090402 (2014).
22. G. A. Howland, J. Schneeloch, D. J. Lum, and J. C. Howell, "Simultaneous measurement of complementary observables with compressive sensing," *Phys. Rev. Lett.* **112**, 253602 (2014).
23. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.* **95**, 131110 (2009).
24. W. K. Yu, M. F. Li, X. R. Yao, X. F. Liu, L. A. Wu, and G. J. Zhai, "Adaptive compressive ghost imaging based on wavelet trees and sparse representation," *Opt. Express* **22**, 7133–7144 (2014).
25. M. Aßmann and M. Bayer, "Compressive adaptive computational ghost imaging," *Sci. Rep.* **3**, 1545 (2013).
26. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**, 3736–3745 (2006).
27. J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**, 2861–2873 (2010).
28. W. Dong, D. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.* **20**, 1838–1857 (2011).
29. B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Network* **7**, 333–339 (1996).
30. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, 607–609 (1996).
31. H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proceedings of Advances in Neural Information Processing Systems (NIPS, 2006)*, pp. 801–808.
32. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory* **38**, 587–607 (1992).
33. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.* **20**, 33–61 (1998).
34. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.* **3**, 1–122 (2011).
35. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of IEEE International Conference on Computer Vision (IEEE, 2001)*, pp. 416–423.
36. M. S. Lewicki and B. A. Olshausen, "Probabilistic framework for the adaptation and comparison of image codes," *J. Opt. Soc. Am. A* **16**, 1587–1601 (1999).
37. X. Hu, Y. Deng, X. Lin, J. Suo, Q. Dai, C. Barsi, and R. Raskar, "Robust and accurate transient light transport decomposition via convolutional sparse coding," *Opt. Lett.* **39**, 3177–3180 (2014).
38. K. Kavukcuoglu, P. Sermanet, Y. L. Boureau, K. Gregor, M. Mathieu, and Y. L. Cun, "Learning convolutional feature hierarchies for visual recognition," in *Proceedings of Advances in Neural Information Processing Systems (NIPS, 2010)*, pp. 1090–1098.

1. Introduction

Ghost imaging (GI) is a novel imaging technique which non-locally records a scene, and has drawn wide attention over the last two decades. In the scheme of GI, two correlated beams travel along different light paths, one light beam illuminates the scene and is collected by a bucket detector, the other one is directly recorded by a spatially-resolved detector. By correlating the

outputs of these two light paths, one can reconstruct the spatially-resolved image of the scene. The “ghost” emphasizes that only a non-spatially-resolved detector is needed to detect the light interacting with the target scene.

GI has gone through three main development stages in terms of the adopted light sources: quantum entangled photons [1], classical thermal light [2–4], and programmable illuminations [5,6]. From quantum to classical to computational, GI is becoming more and more flexible, and has already been put into various practical applications, such as 3D reconstruction [7], fluorescence imaging [8], optical encryption [9,10], remote sensing [11], looking through atmospheric turbulence [12,13], object tracking [14,15], etc. Among the three, computational ghost imaging successfully transfers the complexity of ghost imaging from the experimental apparatus to computation, and makes it possible to enhance and extend GI with the aid of computational resources.

The reconstruction algorithms of ghost imaging mainly fall into two types: second-order (or higher-order) correlation and compressive sensing. The former one obtains the target image through calculating the second-order (or higher-order) correlations of the bucket detector measurements and the illumination patterns. This method suffers from low reconstruction quality due to limited measurements, although some variants have been proposed to improve the performance [16–19]. Differently, the latter one, compressive ghost imaging (CGI), reconstructs the ghost image based on compressive sensing, which has also been successfully used in other applications, such as phase retrieval [20–22]. Through exploiting the redundancy in the structure of natural images [23], CGI enables ghost imaging from sub-Nyquist measurements and largely reduces the acquisition time. Besides, adaptive methods have been proposed to further decrease the requisite measurements [24,25].

The higher reconstruction quality of CGI is attributed to the utilization of pixel-wise prior knowledge (e.g., minimizing the total variation to enforce the local smoothness) or global prior features (e.g., forcing sparsity of DCT coefficients to ensure the dominance of low frequencies) of natural images. Other than the above priors, there also exist strong priors in image patches. Statistically, the image patches are of low dimension and exhibit simple structures. These structures can be decomposed into several primitives, and different structures may share some primitives. So far, the patch prior has been extensively studied and successfully utilized to achieve state-of-the-art performances in various computer vision tasks [26–28].

In this paper, we propose to unify the patch prior together with the pixel or global prior into the CGI reconstruction framework. As far as we know, this is the first time to utilizing patch prior in natural images for ghost imaging and it is nontrivial, because the collected measurements from the bucket detector encode the information of the holistic scene. Our studies demonstrate that, incorporating the patch prior provides great help to improve the reconstruction quality of the non-spatially resolved imaging technique. The remainder of the paper is organized as follows: Section 2 introduces our modeling and derivation. Section 3 demonstrates the effectiveness of our method on both synthetic and real captured data. Finally, Section 4 makes further discussions and summarizes this work.

2. Method

In general, we introduce a linear indexing operator to map each patch to the whole image and vice versa (as shown in Fig. 1), then the constraint defined locally and globally can be integrated together for an intensive utilization of natural image redundancy. As for the pixel-wise or global prior, we can either minimize the total variation or enforce the sparsity of DCT coefficients. As for the patch prior, we represent the image patches by a composition of several primitives, which depict the elementary local structures of natural images, and enforce the sparsity of the representation coefficients. Since both the total variation minimization and coefficient sparse-

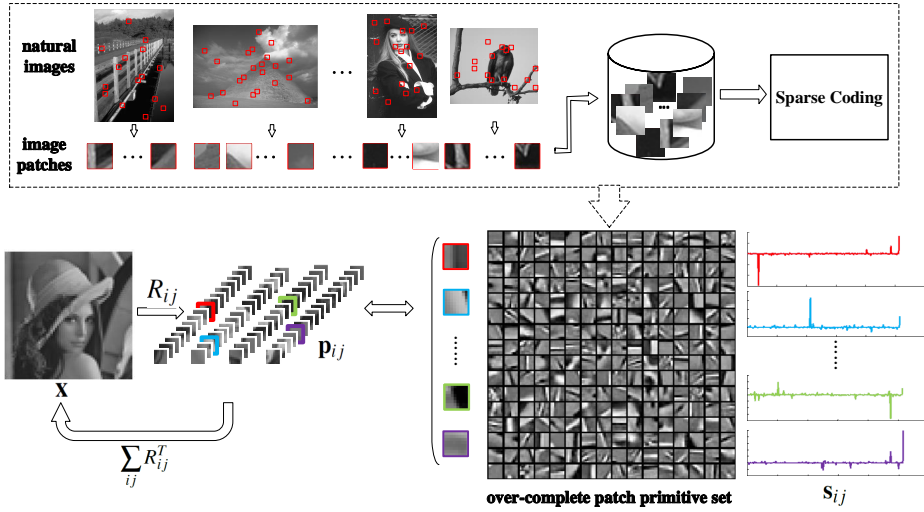


Fig. 1. Schematic illustration of our model. The upper part (framed with dashed box) depicts the learning process of patch primitive set. For a given image, each patch \mathbf{p}_{ij} can be extracted from the image \mathbf{x} by \mathbf{R}_{ij} , and represented with sparse coefficients \mathbf{s}_{ij} over the learned over-complete patch primitive set. Inversely, with the patch-to-image mapping $\{\mathbf{R}_{ij}^T\}$, we could reconstruct the whole image using the over-complete patch primitive set and the corresponding sparsity coefficients $\{\mathbf{s}_{ij}\}$.

ness (either of local patch or the holistic image) can be achieved by minimizing a l_1 norm, we can unify different constraints within a convex optimization framework to reconstruct the target scene.

The image patches represent local regions of a natural image. Usually, the patches are defined in terms of fixed-size blocks (e.g. $k \times k$ pixels), which is much smaller than the original image size, as illustrated in the upper part of Fig. 1. Statistical studies suggests that natural images contain characteristic structures that set them apart from random images (i.e. the pixel intensities are random) [29]. Therefore, characterizing the structures of natural images, and formulating the properties effectively may lend insights into the recovery of natural images [30]. Researches have shown that, one can apply the sparse coding algorithm to a large image patch set to learn a primitive set indicating the basic structures of natural image patches [30,31], and each patch can be represented by sparse linear combination of these primitives. The subfigure framed with dashed box in Fig. 1 gives an schematic illustration of the process of learning the patch primitive set.

Concretely, given a large number of natural image patches $\{\mathbf{p}^1, \dots, \mathbf{p}^U\} \in \mathbb{R}^{k \times k}$ which are randomly cropped from natural images databases as input, the goal of sparse coding is to find the patch primitive set $\{\mathbf{d}^1, \dots, \mathbf{d}^V\} \in \mathbb{R}^{k \times k}$ and the sparse vectors of weight $\{\mathbf{s}^1, \dots, \mathbf{s}^U\} \in \mathbb{R}^{V \times 1}$, such that for each patch $\mathbf{p}^u \approx \sum_{v=1}^V \mathbf{s}_v^u \mathbf{d}^v$, where \mathbf{s}_v^u denotes the v th element of \mathbf{s}^u , i.e. the representation coefficient upon patch primitive \mathbf{d}^v of patch \mathbf{p}^u . The sparse coding problem is formulated as:

$$\arg \min_{\{\mathbf{s}^u\}, \{\mathbf{d}^v\}} \sum_{u=1}^U \|\mathbf{p}^u - \sum_{v=1}^V \mathbf{s}_v^u \mathbf{d}^v\|_2^2 + \beta \sum_{u=1}^U \|\mathbf{s}^u\|_1. \quad (1)$$

Here the first term measures how well the patch primitives represent the image patches, and the second term adopts the L1 norm to enforce the sparsity of the representation coefficients. The parameter β is a positive constant balancing the importance of two terms. We optimize the

objective function over a large database of natural image patches to obtain an over-complete patch primitive set, which is applicable to sparsely represent various natural image patches. Note that the patch primitive set is usually over-complete (i.e. $V > k^2$). This point has been emphasized in [32], which states that the local structures described by the primitives may occur at continuum positions and scales, and over-complete patch primitives could allow for smooth interpolation along this continuum. In other words, over-completeness results in representation accuracy and coefficients sparseness tolerating small translations and scaling of local structures, and thus ensures high flexibility to diverse target images. Overall, the over-completeness of the dictionary ensures the coefficient sparsity an effective prior for general image patches.

We also visualize the learned patch primitive set in Fig. 1, with each entry representing a primitive. Further looking into the details of the primitive set, we could find that the set of primitives describe visually elementary features, such as oriented and translated edges, curves, corners, blobs etc. For each patch from the target scene, we can represent it with a small number out of the whole primitive set, as plotted in Fig. 1. This patch prior has already been successfully incorporated to improve the performance of the classical computer vision tasks, such as denoising [26], superresolution [27] and deblurring [28], etc. In this paper, we propose to incorporate this patch prior into the CGI to further enhance its performance.

To formulate the problem more concisely in matrix form, we let $\mathbf{D} \in \mathbb{R}^{k^2 \times V}$ denote the over-complete primitive set (each column as a vectorized form of a patch primitive), and term it as a over-complete dictionary. Given such a dictionary \mathbf{D} , each patch could be represented as

$$\mathbf{p} = \mathbf{D}\mathbf{s}, \quad (2)$$

in which $\mathbf{p} \in \mathbb{R}^{k^2 \times 1}$ denotes the vectorized image patch, and \mathbf{s} is the representation coefficient of \mathbf{p} . We introduce the over-complete dictionary and the reconstruction coefficients into the CGI reconstruction as in [26]:

$$\mathbf{R}_{ij}\mathbf{x} = \mathbf{p}_{ij}, \quad (3)$$

where \mathbf{x} denotes the whole image, $\mathbf{R}_{ij}(\mathbf{x})$ is the linear image-to-patch mapping, which extracts the patch with its top left pixel locating at (i, j) in image \mathbf{x} . Correspondingly, we denote $\mathbf{R}_{ij}^T(\mathbf{p}_{ij})$ as the inverse mapping, i.e. the patch-to-image operator, and the image \mathbf{x} can be obtained by tiling all the non-overlapping patches, i.e.

$$\mathbf{x} = \sum_{ij} \mathbf{R}_{ij}^T(\mathbf{p}_{ij}). \quad (4)$$

Based on above notations, we can introduce the patch prior into the reconstruction algorithm of conventional CGI (CCGI):

$$\begin{aligned} \arg \min_{\mathbf{s}_{ij}} \quad & \|\Psi\mathbf{x}\|_1 + \lambda \sum_{ij} \|\mathbf{s}_{ij}\|_1 \\ s.t. \quad & \|\Phi\mathbf{x} - \mathbf{y}\|_2^2 \leq \varepsilon, \\ & \mathbf{x} = \sum_{ij} \mathbf{R}_{ij}^T \mathbf{p}_{ij}, \\ & \mathbf{p}_{ij} = \mathbf{D}\mathbf{s}_{ij}. \end{aligned} \quad (5)$$

The first objective term is used to impose the image prior defined over the whole image (e.g, enforcing a small total variation or sparse DCT coefficients), with Ψ being the transform matrix applied on the whole image. The second term imposes the local prior that each image patch is of sparse representation coefficients upon the over-complete dictionary. We use a weighting factor λ to balance these two objective terms. The first inequality constraint describes the data

fidelity. Here Φ is the measurement matrix, with each row denoting a vectorized illumination pattern, \mathbf{y} is the measurements, and variable ε is introduced for robustness to the measurement noise. The second constraint builds a mapping between the image patches and the whole image, and facilitates formulating local and global priors in a unified framework. The third constraint comes from the representation of image patches defined in Eq. (2). Comparatively, CCGI [23] recovers ghost images through minimizing the first objective term under the first inequality constraint. Since utilizing only the whole image prior is prone to smoothing out the thin structures in the target image, we introduce patch prior (the second objective term and two additional equality constraints) to preserve the local details better. In this paper, we name the extended optimization in Eq. (5) primitive-driven CGI (PCGI).

Removing the intermediate variables \mathbf{x} and \mathbf{p}_{ij} in Eq. (5), we could get

$$\begin{aligned} \arg \min_{\mathbf{s}_{ij}} \quad & \|\Psi \sum_{ij} \mathbf{R}_{ij}^T (\mathbf{D}\mathbf{s}_{ij})\|_1 + \lambda \sum_{ij} \|\mathbf{s}_{ij}\|_1 \\ \text{s.t.} \quad & \|\Phi \sum_{ij} \mathbf{R}_{ij}^T (\mathbf{D}\mathbf{s}_{ij}) - \mathbf{y}\|_2^2 \leq \varepsilon. \end{aligned} \quad (6)$$

We could rewrite the problem in Eq. (6) in a similar way to [33] as

$$\arg \min_{\mathbf{s}_{ij}} \|\Psi \sum_{ij} \mathbf{R}_{ij}^T (\mathbf{D}\mathbf{s}_{ij})\|_1 + \lambda \sum_{ij} \|\mathbf{s}_{ij}\|_1 + \frac{\mu}{2} \|\Phi \sum_{ij} \mathbf{R}_{ij}^T (\mathbf{D}\mathbf{s}_{ij}) - \mathbf{y}\|_2^2, \quad (7)$$

where μ is the penalty parameter of the measurement noise.

To solve the problem, we denote m, n as the number of patches along two dimensions of the image and i, j as the corresponding indices, and then reformulate the problem as

$$\arg \min_{\tilde{\mathbf{s}}} \|\Psi \tilde{\mathbf{R}}^T \tilde{\mathbf{D}} \tilde{\mathbf{s}}\|_1 + \lambda \|\tilde{\mathbf{s}}\|_1 + \frac{\mu}{2} \|\Phi \tilde{\mathbf{R}}^T \tilde{\mathbf{D}} \tilde{\mathbf{s}} - \mathbf{y}\|_2^2. \quad (8)$$

Here $\tilde{\mathbf{R}}^T = \text{blkdiag}(\{\mathbf{R}_{ij}\})$, $\tilde{\mathbf{D}}^T = \text{blkdiag}(\{\mathbf{D}_{ij} | \forall ij \mathbf{D}_{ij} = \mathbf{D}\})$ are two block diagonal matrices, with $\text{blkdiag}(\cdot)$ denoting diagonal concatenation of the input entries; $\tilde{\mathbf{s}} = [\mathbf{s}_{11}^T, \dots, \mathbf{s}_{mn}^T]^T$ is a vector composed by concatenating the reconstruction coefficients of all $m \times n$ image patches.

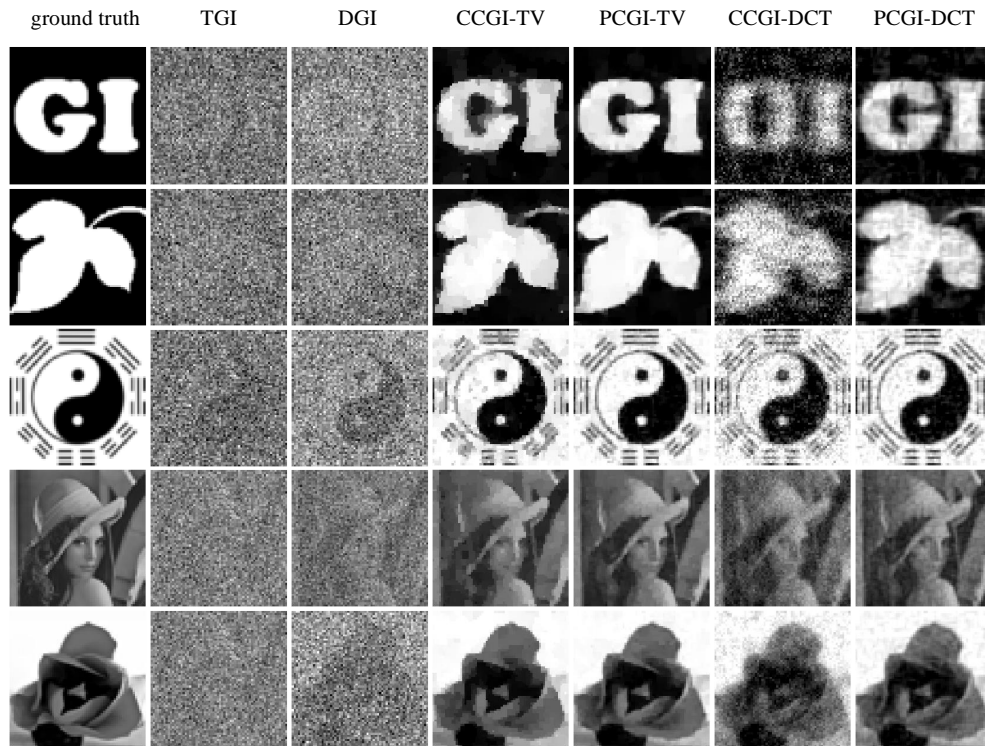
By simple variable substitution, Eq. (8) could be further rewritten into

$$\begin{aligned} \arg \min_{\tilde{\mathbf{s}}} \quad & \|\mathbf{w}\|_1 + \lambda \|\tilde{\mathbf{s}}\|_1 + \frac{\mu}{2} \|\Phi \tilde{\mathbf{R}}^T \tilde{\mathbf{D}} \tilde{\mathbf{s}} - \mathbf{y}\|_2^2 \\ \text{s.t.} \quad & \mathbf{w} = \Psi \tilde{\mathbf{R}}^T \tilde{\mathbf{D}} \tilde{\mathbf{s}}. \end{aligned} \quad (9)$$

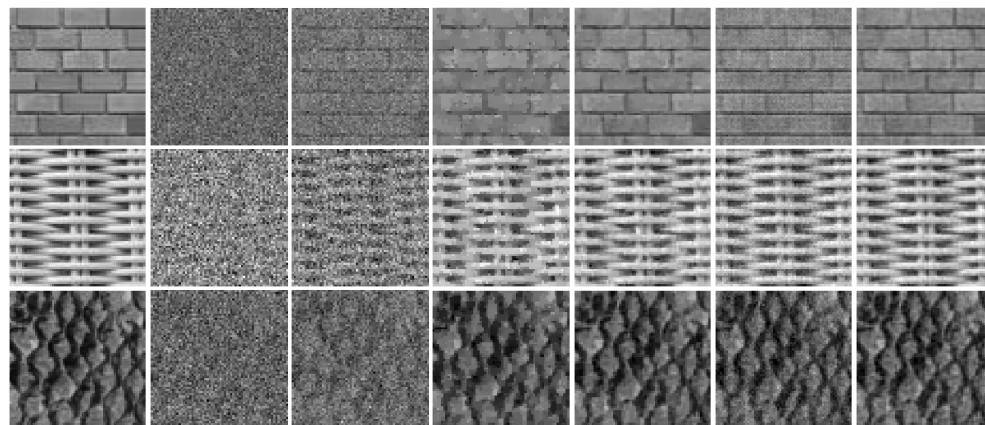
Eq. (9) falls into a typical convex optimization, and we resort to the alternating direction method of multipliers algorithm [34] to solve the model. The final reconstructed image could be obtained from the optimum $\tilde{\mathbf{s}}^*$ according to Eq. (4).

3. Experiments on simulation data

To demonstrate the performance of our method, we conduct a series of numerical simulations upon natural images. In implementation, the image size in this experiment are all 64×64 pixels. The measurements are generated by calculating their inner products with different random binary patterns. We set the patch size to be 8×8 pixels empirically. The over-complete dictionary is trained from 300 natural images from the Berkeley Segmentation Data Set 300 (BSDS300) [35]. For each natural image, we choose 250 patches at randomly chosen positions, and in sum we use 750,000 patches. Each patch is normalized to zero mean before provided to the efficient sparse coding algorithm [31]. As shown in Fig. 1, there are in all 256 primitives. As discussed before, the number of primitives should be bigger than the dimension of the image



(a)



(b)

Fig. 2. Performance comparison on simulated data. The 1st column displays the ground truth images. The 2nd and the 3rd column is respectively the reconstruction results of TGI and DGI. The 4th vs. 5rd, and 6th vs. 7th columns compare the reconstruction results of CCGI and PCGI. (a) Images without periodic textures: letters 'GI' (SSR=0.08), 'Leaf' (SSR=0.10), 'Eight-trigrams' (SSR= 0.35), 'Lena' (SSR= 0.35), 'Flower' (SSR= 0.25). (b) Images with periodic textures: 'Brickwall' (SSR= 0.45), 'Wickerwork' (SSR= 0.5), 'Fishscale' (SSR= 0.45).

patch to ensure the over-completeness, i.e., the sparsity of the patches' reconstruction coefficients. Although theoretically a larger primitive set leads to better reconstruction, adopting too large a primitive set would pose heavy computational load on the reconstruction algorithm. To balance the over-completeness and computation load, we set the primitive set size as 256, the same as [26], which was demonstrated empirically to achieve good performance for general natural images. Accounting for the de-mean normalization of the training patches, we add a DC patch (all the entries of the 8×8 primitive is set to be 1) to the trained dictionary. By vectorizing each primitive as a column vector, we could get the over-complete dictionary \mathbf{D} (64×257).

For a better evaluation of the advantages of introducing the patch prior, we compare the reconstruction results with and without it. Without loss of generalization, we use two widely adopted Ψ s—gradient operator and 2D-DCT domain transform matrix, as in [11, 23, 24]. To express compactly, we denote two transform matrices as Ψ_{tv} and Ψ_{dct} , respectively. As for the two parameters in Eq. (9), we found that experimentally, the algorithm is insensitive to their values as long as they fall into an appropriate but reasonably wide range. Throughout the experiments, we use the same parameter setting: $\lambda = 0.1$, $\beta = 0.4$ and $\mu = 1000$. Besides, for a comprehensive evaluation of our approach, other than CCGI, we also compare our method with the correlation-based computational GI methods, including traditional GI (TGI) [6], differential GI (DGI) [16], normalized GI (NGI) [17] and iterative GI (IGI) [18]. Considering that our experiments were conducted with the sub-Nyquist sampling ratio (SSR), i.e., the number of illumination patterns is smaller than the resolution of the target image, in which case NGI and IGI exhibit similar performance to DGI (slightly better than TGI), we choose to show only the results of TGI and DGI here.

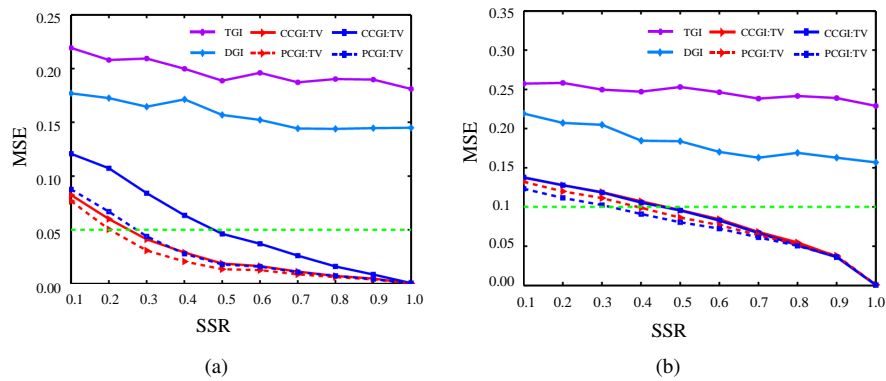


Fig. 3. Quantitative performance comparison with respect to SSRs among six different algorithms referred in Fig. 2. (a) 'Lena', (b) 'Wickerwork'.

To test the performance of our approach at different SSRs, for each scene we reconstruct the images with SSR ranging from 0.1 to 1 with an interval of 0.1. Here we choose to display the reconstruction results at the SSR with the most prominent performance difference in Fig. 2. Comparatively, CS-based methods (i.e. PCGI-TV, PCGI-DCT, CCGI-TV, CCGI-DCT) perform much better than the correlation-based methods (i.e. TGI, DGI). The suffixes 'TV' and 'DCT' respectively denotes using Ψ_{tv} and Ψ_{dct} as the transform matrix Ψ in Eq. (5). The large difference in reconstruction performance comes from their different reconstruction mechanisms: The correlation-based methods rely only on the statistical properties of the cross-correlation matrix of the illumination patterns, thus requires large number of measurements, usually more than the size of the target image. On the contrary, CS-based methods incorporate priors additionally and thus are of good quality even at sub-Nyquist ratio. Comparing PCGI methods (i.e.

PCGI-TV, PCGI-DCT) with CCGI methods (i.e. CCGI-TV, CCGI-DCT), we could find that the latter demonstrates better reconstruction, especially in the areas with edges (e.g. the boundary of the letters 'GI', the blade of 'Leaf', the hat brim of 'Lena', the petal of the 'Flower', the gaps among the fish sales), lines (e.g. the petiole of the 'Leaf', the lines around the 'Eight-trigrams', the thin grooves of the 'Brickwall', the wicker of the 'Wickerwork'), and blobs (e.g. the dots in the 'Eight-trigrams', the stigma of the 'Flower'). The performance gain of PCGI is mainly attributed to the patch primitive prior. Observing the learned patch primitive set in Fig. 1, we can find that edges, lines and small blobs are typical patch primitives, thus image patches exhibiting such structures are more likely to be composed by a sparse combination of the learned primitives. Therefore, introducing patch prior would improve the performance by enhancing the high frequencies that tend to be smoothed out by imposing prior defined on the whole image—minimizing total variation (CCGI-TV) or favouring the dominance of low frequencies (CCGI-DCT).

Besides, we can find experimentally that the performances at different settings are slightly different for the images in Figs. 2(a) and 2(b). For images without periodic textures in Fig. 2(a), the reconstruction using total variation prior exhibits higher quality than using DCT prior. The advantage of introducing patch prior is more distinct when using Ψ_{dct} . On the contrary, in Fig. 2(b), the algorithm with DCT prior works better in the cases with rich periodic textures. The patch prior exhibits a more marked improvement for CGI with total variation prior. The varying applicability to different cases can help choosing different scene specific algorithms.

For quantitative evaluation, we adopt mean square error (MSE) as the evaluation metric and compare the reconstruction with respect to SSR among four different algorithms. In Figs. 3 (a) and 3(b), we respectively plot the MSE comparison of 'Lena' and 'Wickerwork', as two representative examples for the images in Figs. 2(a) and 2(b). As the number of measurements is far from being sufficient for decent correlation-based reconstruction, the performances of both TGI and DGI are apparently inferior to those of CCGI and PCGI. Further comparing CCGI and PCGI, one can clearly see that reconstruction of PCGI is better than CCGI, especially at low SSRs from 0.2 to 0.5. This implies that our method can obtain the same reconstruction quality with much less measurements. As in Fig. 3(a), to retrieve an image with the same quality (MSE= 0.05), our method could reduce the measurements by 5.7% and 22% for Ψ_{tv} and Ψ_{dct} transform, respectively. Similarly, in Fig. 3(b), for image 'Wickerwork', with abundant periodic texture, our method can reduce 5.3% and 12.5% measurements to achieve reconstruction with MSE= 0.1. In all, these results demonstrate the contribution of our approach in enhancing the quality, and thus reducing the requisite acquisition time of CGI.

Considering the inevitable sensor noise in an imaging system, we further test the reconstruction performance at varying noise levels. We simulate imaging noise by superimposing additive Gaussian white noise with the signal to noise ratio (SNR) from 10dB to 70dB. In this experiment, we set SSR to be 0.25 and provide the reconstruction results of six different methods with MSE as the evaluation metric. We choose to show the comparison result of 'Lena' and 'Wickerwork' in Fig. 2. As plotted in Fig. 4, the reconstruction quality of all methods increases monotonously as the noise level decreases and become stable when the signal to noise is higher than 40dB. As can be seen, our method performs much better than the other methods. The comparison between using different Ψ s is consistent with the comparison discussed above.

4. Experiments on real captured data

Then, we experimentally demonstrate our method on the data captured by the prototype system. The scheme of our experimental setup for computational GI is exhibited in Fig. 5. The illumination module is built by hacking a commercial projector. Firstly, the light emitted from the halogen lamp is successively converged by a condenser lens, collimated by an optical integrator

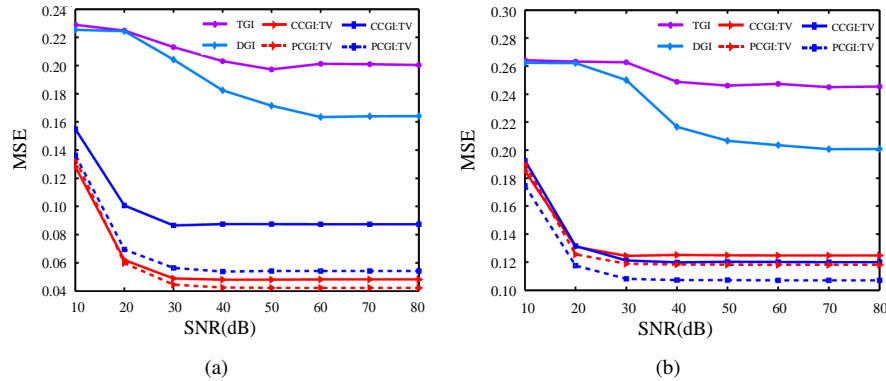


Fig. 4. Comparison of robustness to sensor noise among six algorithms. (a) 'Lena', (b) 'Wickerwork'.

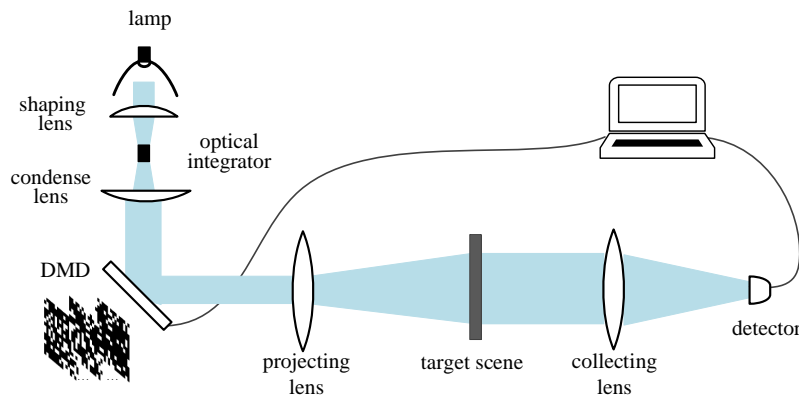


Fig. 5. The schematic diagram of CGI. Light source: high pressure Mercury lamp (Philips, 200w). DMD: Texas Instrument DLP® Discovery™4100, .7XGA. Scene: a transmissive film (34mm×34mm). Detector: Thorlabs DET100 Silicon photodiode (integration time: 0.625ns).

and then adjusted by a shaping lens. Then, the light is modulated by the digital micro-mirror device (DMD) to generate random patterned illuminations. After beam expansion with a projector lens, the patterned illumination interacts with the scene and the outgoing photons are collected by a bucket detector after going through a converging lens. The measurements are collected by sampling the outputs of the detector using a 14bit acquisition board ART PCI8514, which performs analog-to-digital (AD) signal conversion and then transmits the data to the computer for follow-up reconstruction. The illumination pattern on the DMD is controlled by the computer, and the target image is retrieved from the illumination patterns and the bucket measurements.

In our CGI setup, there are several factors that might influence the reconstruction performance. We investigate the influences from instrument specifications and adopt corresponding strategies to avoid the performance degeneration.

At the illumination side, since we use binary (i.e., $\{0, 1\}$) patterns, there is no quantization error during the DMD modulation. The AD at the acquisition board would introduce quantization error. Here we adopt a 14-bit digitalization depth, and the fluctuation range of the single pixel detector is 0-5000mV, so the resolution of AD is 0.3053mV. In our experiments, the measure-

ment range among the patterns are respectively $825 \pm 67.6425\text{mV}$ and $1103 \pm 108.2650\text{mV}$ with respect to the letters 'GHOST IMAGE' and 'Ghost'. So the quantization error occurs at the acquisition board can also be ignored.

Theoretically, one can get stair shaped measurement transition, but the elapse caused by pattern transition of the illumination and integration of the detector would smooth the transition edges. In our experiment, the illumination frequency is set to be 100Hz. We set the acquisition frequency at 100 times of the illumination, i.e., 10000Hz. So the sampling period $100\mu\text{s}$ and there are 100 measurements for each pattern. Because the integration time is 0.625ns and the DMD transition takes $12\mu\text{s}$, we discarded 5 points in the beginning and ending of the 100 measurements to avoid the influence of the elapse caused by either illumination or detector. In implementation, we take average over the 90 left samples as the measurement for each pattern.

The sampling frequency can also influence the reconstruction performance, so we specially compare the performances at different sampling frequencies (ranging from 1000Hz to 100000Hz). As in Fig. 6, as the sampling frequency increases, the fidelity of reconstruction improves. This trend is reasonable, since denser sampling would suppress the sensor noise better with respect to each pattern. Similar to the trend on simulated data, display in Fig. 5, the reconstruction quality stop improving when the sampling frequency exceeds 10000Hz. This is mainly due to that our method is robust to small noise, thus the reconstruction result will not be distinctly improved by further raising the SNR of the measurements. Therefore, in our experiment, we set the sampling frequency to be 10000Hz.

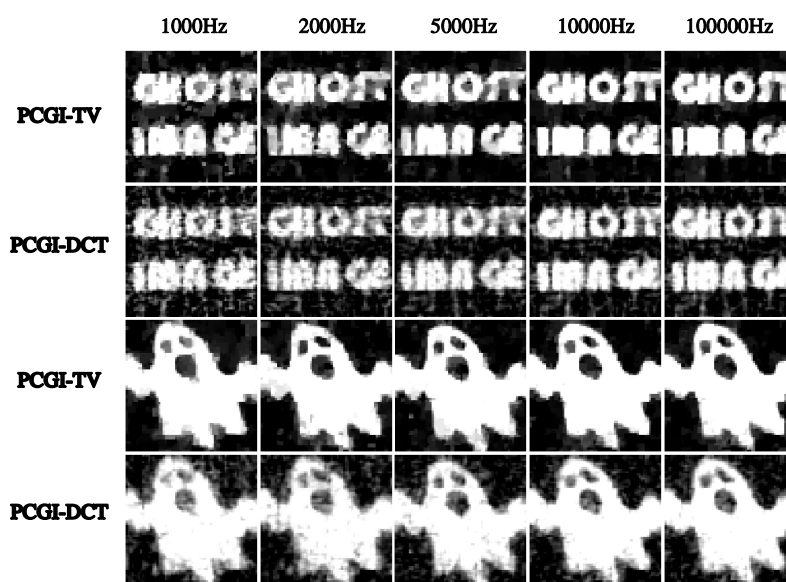


Fig. 6. The reconstructions of 'GHOST IMAGE' and 'Ghost' from the measurements of different sampling frequencies.

In our experiments, we reconstruct the image 'Ghost' and the uppercase alphabet 'GHOST IMAGE'. The spatial resolution is 64×64 pixels. We collect 614 measurements ($\text{SSR} = 0.15$) for each image. As can be seen in Fig. 7, the comparison between our approach and conventional CGI exhibit a similar trend with that on simulation. Although the reconstructions are a little bit corrupted due to the noise during sampling, by introducing the patch prior, our method could still show its superiority in reconstructing the local details of the image. For example, we can obtain sharper edges of the alphabet 'GHOST IMAGE' and cleaner outline of the 'Ghost'.

Here we make further effort upon quantitative performance comparison. Although we have the image of the target scene used in film printing, taking the film image as ground truth suffers from misalignment with the reconstruction. Hence, we choose to use the high quality reconstruction result of PCGI-TV with a large SSR (i.e. SSR= 0.5) as a pseudo-ground truth (PGT), as shown in the first column of Fig. 7. Here we still use MSE as the evaluation metric. Corresponding to the reconstructed images from 2nd to 7th column in Fig. 7, the MSEs of ‘GHOST IMAGE’ are 0.460, 0.447, 0.091, 0.068, 0.149, 0.086, and those of ‘Ghost’ are 0.464, 0.441, 0.078, 0.054, 0.177, 0.089. The performance ranking is consistent to the visual comparison, and the effectiveness of the proposed approach is further validated.

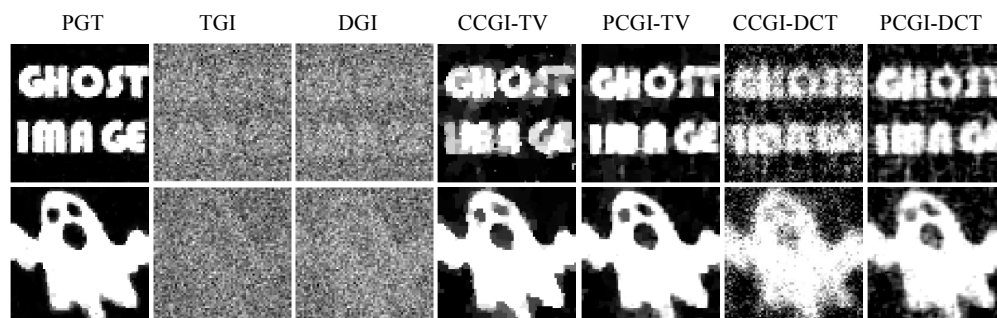


Fig. 7. The reconstructions from the measurements captured by our prototype. The 1st column displays the PGT, the 2nd and 3rd columns show results of TGI and DGI, and the 4th to 7th column shows the reconstruction by CCGI-TV, PCGI-TV, CCGI-DCT, PCGI-DCT, respectively.

5. Conclusions and discussions

In summary, this paper proposes to introduce the patch prior of natural images into the computational ghost imaging framework. Experiments show that our approach largely raises the reconstruction quality of ghost imaging. The superiority to conventional CGI is mainly attributed to utilizing the patch prior, which enforces each patch to be a sparse linear combination of primitives learned from a natural image database.

Mathematically, the dimension of the optimization variable increases after introducing the patch primitives, so the proposed approach is of higher computational complexity than conventional CGI. Fortunately, the related heavy calculations, such as the matrix inversion to get greedy solution at each iteration, is scene independent and can be calculated off-line before collecting measurements and conducting reconstruction. Therefore, introducing patch prior does not lead heavy computation and can serve as a feasible solution for improving quality of ghost imaging.

One promising extension of our method is to perform adaptive dictionary learning to take advantage of the high flexibility of sparse coding, i.e. if we know the type of the target scene, we could choose the same type of training image patches to learn a specific over-complete dictionary for better reconstruction [36]. Convolutional sparse model [37, 38] can also be utilized to incorporate the patch prior into the framework of CGI and has the potential of higher performance. Besides, we also plan to introduce self similarity existing in the target image to reduce the required measurements further. Then, dynamic ghost imaging tends to be feasible and will broaden the practical applications of CGI.

Acknowledgments

This work was supported by the projects of National Science Foundation of China (Nos. 61327902 and 61120106003), NSF award 1115680. The research is also funded by Beijing Key Laboratory of Multi-dimension & Multi-scale Computational Photography (MMCP), Tsinghua University.