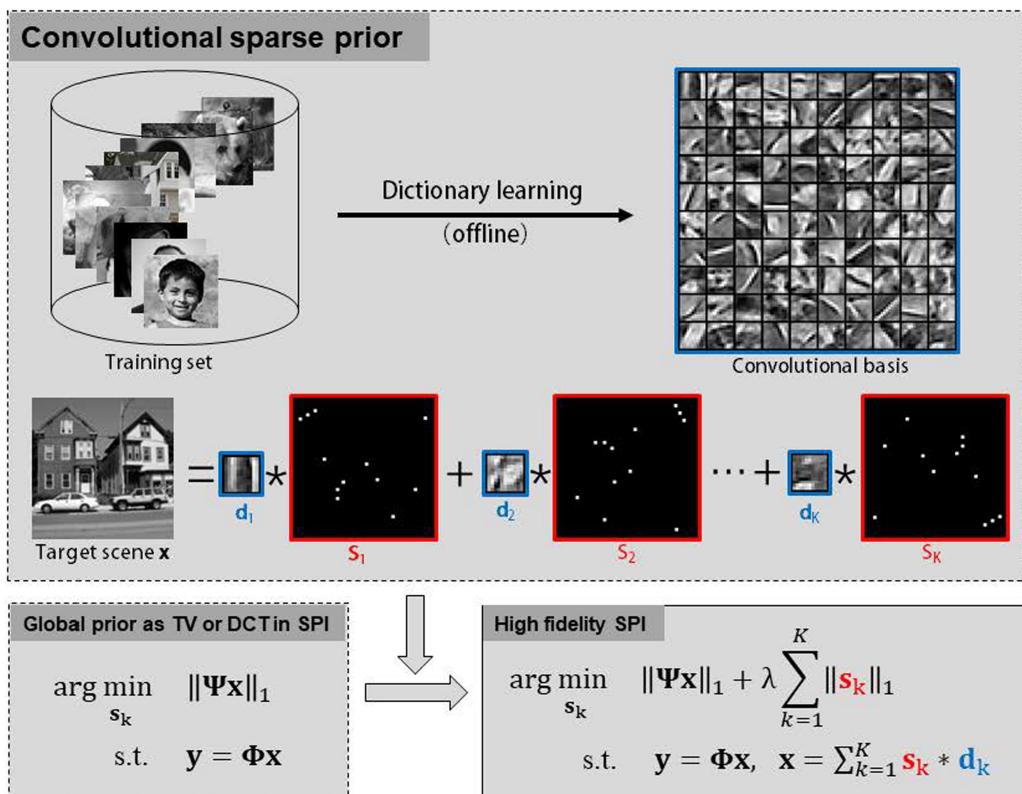


High Fidelity Single-Pixel Imaging

Volume 11, Number 2, April 2019

Chao Deng
 Xuemei Hu
 Xiaoxu Li
 Jinli Suo
 Zhili Zhang
 Qionghai Dai, *Senior Member, IEEE*



DOI: 10.1109/JPHOT.2019.2900549

1943-0655 © 2019 IEEE

High Fidelity Single-Pixel Imaging

Chao Deng ^{1,2} **Xuemei Hu**, ³ **Xiaoxu Li**, ¹ **Jinli Suo** ¹ **Zhili Zhang**, ²
and **Qionghai Dai** ¹ **Senior Member, IEEE**

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

²High-Tech Institute of Xi'an, Xi'an 710025, China

³School of Electronic Science and Engineering, Nanjing University, Nanjing 210000, China

DOI:10.1109/JPHOT.2019.2900549

1943-0655 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Manuscript received January 15, 2019; revised February 14, 2019; accepted February 17, 2019. Date of publication March 1, 2019; date of current version March 8, 2019. This work was supported by the National Science Foundation of China under Grant 61722110 and Grant 61631009. Corresponding author: Jinli Suo (e-mail: jlsuo@tsinghua.edu.cn).

Abstract: Single-pixel imaging (SPI) is an emerging technique which has attracted wide attention in various fields. However, restricted by the low reconstruction quality and large amount of requisite measurements, SPI's practical application is still in its infancy. Inspired by the fact that natural scenes exhibit unique degenerated structures in the low-dimensional subspace, we propose to take advantage of such local prior via convolutional sparse coding to implement high fidelity SPI. Specifically, we can represent the target scene via convolving with a set of statistically learned kernels, with the convolution coefficient matrix being sparse. We introduce the above local prior into conventional SPI framework to promote the final reconstruction quality. Experiments both on synthetic data and real captured data demonstrate that our method can achieve better reconstruction from the same measurements, or reduce the number of required measurements for the same reconstruction quality.

Index Terms: Single-pixel imaging, high fidelity, sparse coding, local prior.

1. Introduction

Single-Pixel imaging (SPI) is an emerging imaging scheme capturing the target scene using a photodiode. Specifically, the object can be reconstructed by correlating the outputs of two beams: one interacts with the target object and is recorded by a bucket detector, and the other is directly collected by a spatially resolved detector. At the beginning, SPI was considered as a unique imaging mechanism of entangled photons in quantum physics [1], which was referred as ghost imaging (GI). Soon after, such configuration is also proved to be applicable for classic thermal light sources [2]–[4]. Replacing the spatially resolved detector in the reference arm with programmable illumination makes SPI more flexible and ready to be put into practical applications, such as fluorescence imaging [5], remote sensing [6], 3D reconstruction [7], optical encryption [8], [9], and object tracking [10], [11].

So far, three typical categories of reconstruction algorithms for SPI have been proposed: linear correlation [12]–[15], the alternating projection (AP) [16]–[18] and compressive sensing (CS) based reconstruction [19], [20]. The correlation-based methods restore the target scene through second order or higher order correlations, which compute fast but suffer from low reconstruction quality. The AP algorithm incorporates the constraints from the patterned illumination and correlated measurements in spatial and the Fourier domain alternatively. It typically iterates 100~200 rounds (each round includes iterations incorporating information from all the measurements) until the final

convergence. In contrast, the compressed sensing algorithm in SPI introduces the image priors into the under-determined linear system to reduce the solution space, and thus needs much fewer measurements and of faster convergence.

In terms of image priors, researchers generally uses global and local priors. Two widely used global priors for the target images are representation sparsity and small total variation (TV) [21]. The former states that natural images can be sparsely represented in some orthogonal basis such as DCT and wavelet [22], [23], and the latter considers the high frequency components are statistically of low probability. Besides, the natural images locally exhibit unique degenerate structures, which can be sparsely denoted in the low-dimensional self-characteristic subspace. For example, image statistics suggest that the image patches can be well represented by a small subset of an over-complete dictionary [24]–[27], which is termed as the sparse coding (SC). Local prior as such in sparse coding has been widely used and achieved state-of-the-art performances in computer vision tasks, e.g., denoising, deblurring, inpainting, super resolution imaging and machine learning. Generally, patch-based window sliding is the most widely used reconstruction scheme. Each patch can be sparsely represented and the whole image is reconstructed by averaging these patches. Patch-based SC considers overlapped patches separately and average them for final reconstruction. Although the patch-based method can achieve high-quality representation, two arising limitations suppress the final reconstruction. First, since the small variance of intrinsic structures (such as shift or rotation) is indistinguishable due to the independent mechanism in patch-based SC, the learned dictionary is highly redundant. Second, the overlap-averaging mechanism leads to inconsistency among overlapped patches [28]. Moreover, the huge computational load in optimization constraints the feasibility in high resolution imaging. For example, [27] can only handle the 64×64 pixel images.

As opposed to the patched-based SC, the convolutional sparse coding (CSC) decomposes the whole image into some sparse feature maps, and thus avoid the prior consistency of learned atoms and inconsistency among overlapping patches. In this paper, we propose to impose local prior via convolutional sparse coding under the compressive sensing framework for high fidelity single-pixel imaging. Through convolution operation, the CSC can handle images with higher pixel resolution than patched-based SC. Our approach can also greatly improve the reconstruction quality in non-spatially resolved imaging technique compared with current state-of-the-art methods.

2. Method

To fully exploit the image prior both globally and locally, we incorporate these two priors together for high fidelity reconstruction. Mathematically, we can learn the kernels of CSC from a large numbers of nature images by following optimization [29], [30]

$$\begin{aligned} \arg \min_{\mathbf{d}, \mathbf{s}} \quad & \sum_{j=1}^J \frac{1}{2} \left\| \mathbf{x}^j - \sum_{k=1}^K \mathbf{d}_k * \mathbf{s}_k^j \right\|_2^2 + \beta \sum_{k=1}^K \|\mathbf{s}_k^j\|_1 \\ \text{s.t.} \quad & \|\mathbf{d}_k\|_2^2 \leq 1 \quad \forall k \in \{1, \dots, K\}, \end{aligned} \quad (1)$$

where each image \mathbf{x}^j can be generated by summing up convolutions of vectorized 2D kernels \mathbf{d}_k and corresponding sparse feature maps \mathbf{s}_k^j . Here \mathbf{x}^j and \mathbf{s}_k^j are both in vectorized form. The coefficient β weights the ℓ_1 penalty, and we set $\beta = 1$ for a proper tradeoff between prior regularity and data fitting accuracy as in [30]. The operation $*$ denotes the 2D convolution defined over the vectorized inputs. We use the fast and flexible algorithm proposed in [30] to obtain the convolutional basis for CSC.

After the kernel learning as shown in Fig. 1, the convolutional basis can be utilized to decompose the target image sparsely. Introducing this local prior into SPI reconstruction, the optimization can

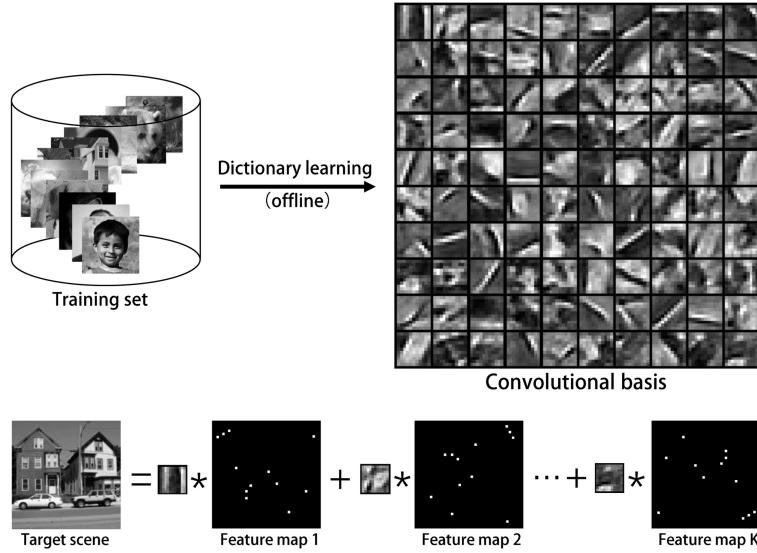


Fig. 1. The overview of our model. The upper part describes the learning process of convolutional sparse coding. With the learned kernels, the target scene can be decomposed as summation of kernels convolved with corresponding sparse feature maps.

be formulated as

$$\begin{aligned} \arg \min_{\mathbf{s}_k} \quad & \|\Psi \mathbf{x}\|_1 + \lambda \sum_{k=1}^K \|\mathbf{s}_k\|_1 \\ \text{s.t.} \quad & \mathbf{y} = \Phi \mathbf{x} \\ & \mathbf{x} = \sum_{k=1}^K \mathbf{s}_k * \mathbf{d}_k. \end{aligned} \quad (2)$$

Here Ψ is the transformation matrix towards the transform domain with sparse representation, such as DCT or TV. Φ is the sampling matrix with each row denoting the vectorized illumination pattern, \mathbf{y} is the single pixel measurement and λ is the balancing weight. The first term in the objective function is the global constraint defined in the DCT or TV domain, and the second term local constraint on the learned dictionary.

For simplicity, we remove the intermediate variable \mathbf{x} and apply simple substitutions to rewrite Eq. 2 as

$$\begin{aligned} \arg \min_{\mathbf{s}_k} \quad & \|\mathbf{v}\|_1 + \lambda \sum_{k=1}^K \|\mathbf{s}_k\|_1 \\ \text{s.t.} \quad & \mathbf{y} = \Phi \cdot \sum_{k=1}^K \mathbf{u}_k \\ & \mathbf{v} = \Psi \cdot \sum_{k=1}^K \mathbf{u}_k \\ & \mathbf{u}_k = \mathbf{s}_k * \mathbf{d}_k. \end{aligned} \quad (3)$$

We solve above optimization by alternating direction method of multipliers (ADMM) algorithm [31], and then the target scene \mathbf{x} can be obtained by the optimum \mathbf{s}_k^* . As for the convolution calculation in Eq. (3), we first transform \mathbf{s}_k and \mathbf{d}_k into the Fourier space, implement Hadamard product, and then



Fig. 2. The reconstruction comparison of three different methods: TV (DCT), TV+SC (DCT+SC), TV+CSC (DCT+CSC), under different compression ratios. We use PSNR and RMSE for quantitative evaluation.

transform back to spatial domain afterwards. This implementation is faster than directly applying convolution in spatial domain, since both Fourier transform and Hadamard product involve only simple matrix multiplication.

3. Experiments on Synthetic Data

To evaluate the performance of our method quantitatively, we conduct a series of simulations. We first learn the convolutional kernels of CSC from 20 randomly chosen images from the database built by the Stanford Vision Lab [32]. Each kernel of the learned dictionary is set to 11×11 pixels, and we have 100 kernels in total (the learned dictionary is shown in Fig. 1). Here we adopt the same dictionary size as in [30] to achieve good performance for general natural scenes. Then we conduct the CSC-based SPI reconstruction using the ADMM optimization. The image size of target reconstruction and modulation patterns are both 128×128 pixels. All the test images are not in the training set. The measurement is generated following the imaging formulation: inner product of the target scene and patterns. Since the CSC is a prior generally compatible with conventional reconstruction, we integrate CSC together with two widely used global priors achieving state-of-the-art performance in SPI — TV minimization and sparsity in DCT domain.

To show the performance of CSC prior, we first conduct a simulation on several images under different compression ratios (i.e., ratio between the number of used patterns and the pixels number

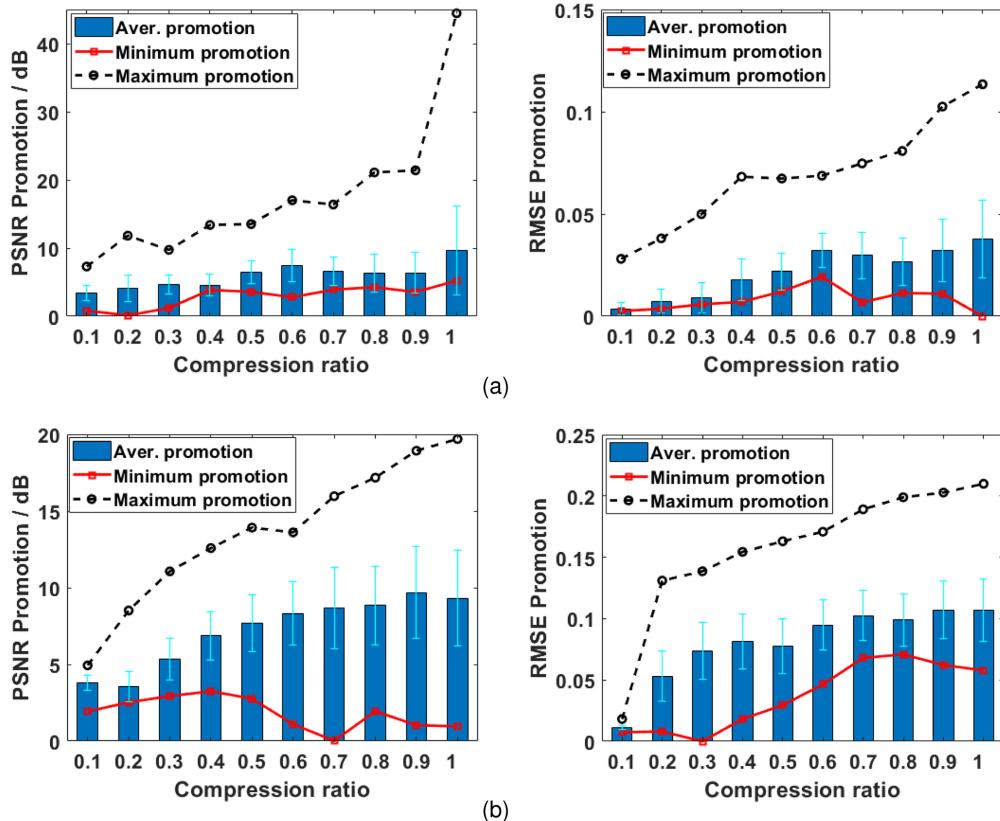


Fig. 3. Promotion of CSC introduced into TV (a) and DCT (b) in terms of PSNR and RMSE under different compression ratios.

in the final reconstruction). We compare the reconstruction of three different methods: conventional global prior (TV, DCT), combining global prior and patch based sparse coding (TV+SC, DCT+SC), and combining global prior with convolutional sparse coding (TV+CSC, DCT+CSC). We vary the sampling ratios at 10%, 15%, 20%, 25%, 30% and 35%. Here for TV+SC or DCT+SC, we use the patch-prior based single pixel imaging algorithm proposed in [27]. The results are shown in Fig. 2. Comparatively, the reconstruction using TV prior exhibits higher quality than DCT prior since most test images are with non-periodic structures. In terms of both peak signal-to-noise ratio (PSNR) and root mean square error (RMSE), introducing CSC can significantly improve the SPI reconstruction. Specifically, the promotion of PSNR is about 3.65 dB~6.30 dB over TV prior, and 5.04 dB~8.67 dB over DCT prior. The improvement in terms of RMSE is also significant, although there exist diversity among different compression ratios and different images. The consistent promotion is attributed to the complementary combination of global and local priors, which further reduces the solution space and improves the reconstruction accuracy. In contrast, the patch-prior based SPI exhibits only limited improvement and at high computation load. For example, to handle an 128×128 image under 10% compression ratio, one requires more than 32.3 GB RAM. The advantages of CSC based local prior comes from the higher sparsity and free of blocking artifacts in computation of the patch overlapping regions. Due to the low efficiency of patch-prior based method in [27], we do not include this method in the following simulations.

To comprehensively evaluating the effectiveness of CSC prior, we investigate the average promotion by CSC on an image set under different compression ratios, ranging from 0.1 to 1.0 with the interval of 0.1. The test data set consists of 12 images from four different groups: person, animal, building and scenery, with 3 images in each group. We plot the average promotion, the error bar

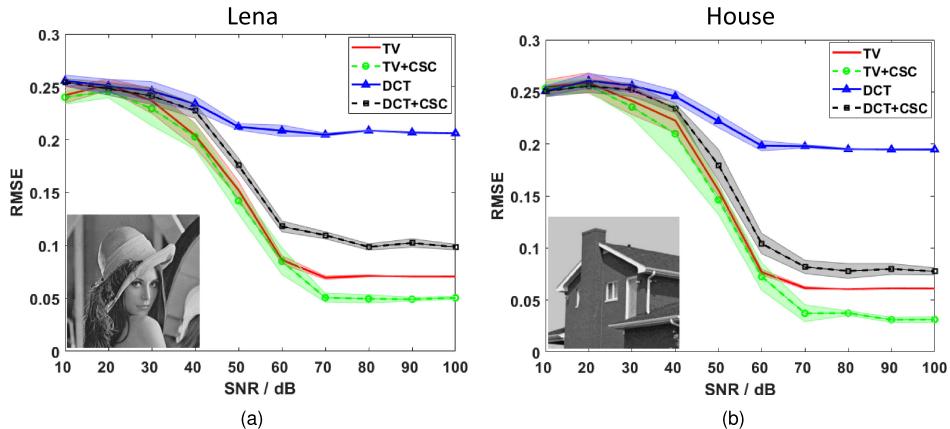


Fig. 4. The performance of four different algorithms at different system noise levels on two exemplar images: (a) "Lena" and (b) "House".

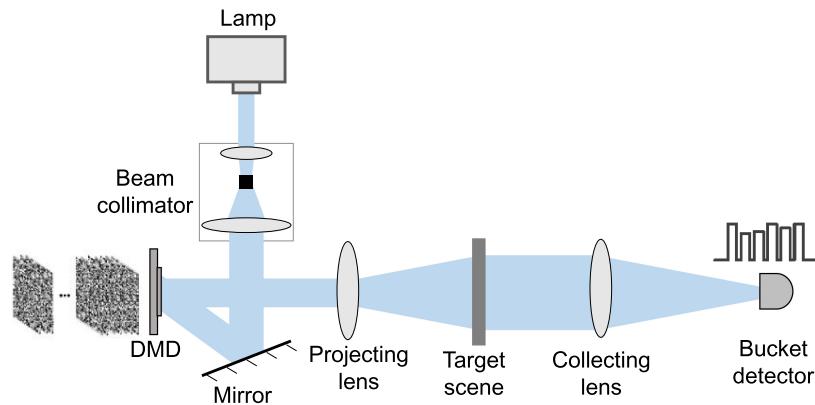


Fig. 5. The schematic diagram of our experimental setup. The high pressure Mercury lamp (Philips, 200 w) is used as the light source, DMD (digital micro-mirror device, Texas Instrument DLP, DiscoveryTM4100, .7XGA) for illumination patterning and a photodiode (Thorlabs DET100 Silicon photodiode, integration time: 0.625 ns) for measurement recording.

quantifying standard deviation, and the extremum in Fig. 3. The result tells that CSC consistently improves the reconstruction under all the settings. The average improvement by integrating CSC prior with TV prior is about 3.4~9.6 dB (PSNR) and 0.004~0.037 (RMSE), while the promotion with respect to DCT is 3.6~9.7 dB (PSNR) and 0.011~0.107 (RMSE). The results show that CSC brings the largest promotion at high sampling ratio both in terms of PSNR and RMSE. Besides, the improvement by introducing CSC exhibit large diversity among test images. In sum, the simulation result in Fig. 3 demonstrates that we can achieve high fidelity image quality in SPI through introducing CSC as prior of local structures in the target scene.

System noise, such as the sensor's readout noise or intensity fluctuation of light source, is inevitable during the actual SPI imaging process, so here we test the noise-robustness of four aforementioned algorithms. We simulate the system noise by superimposing Gaussian white noise onto the measurements at the sensor, with the signal-to-noise ratio (SNR) ranging from 10 dB to 100 dB, and choose RMSE as the evaluation metric. Without loss of generality, we fix the sampling ratio as 0.25 in data synthesis (experimentally the results at different sampling ratios show similar trends). The quantitative performance comparison (i.e., RMSE) among above four algorithms on two classic images — "Lena" and "House" — are shown in Fig. 4. The error for each noise level on the RMSE is determined through Monte Carlo simulation on one hundred samplings. Overall, the reconstruction of all the algorithms improves as the noise level decreases, and becomes stable

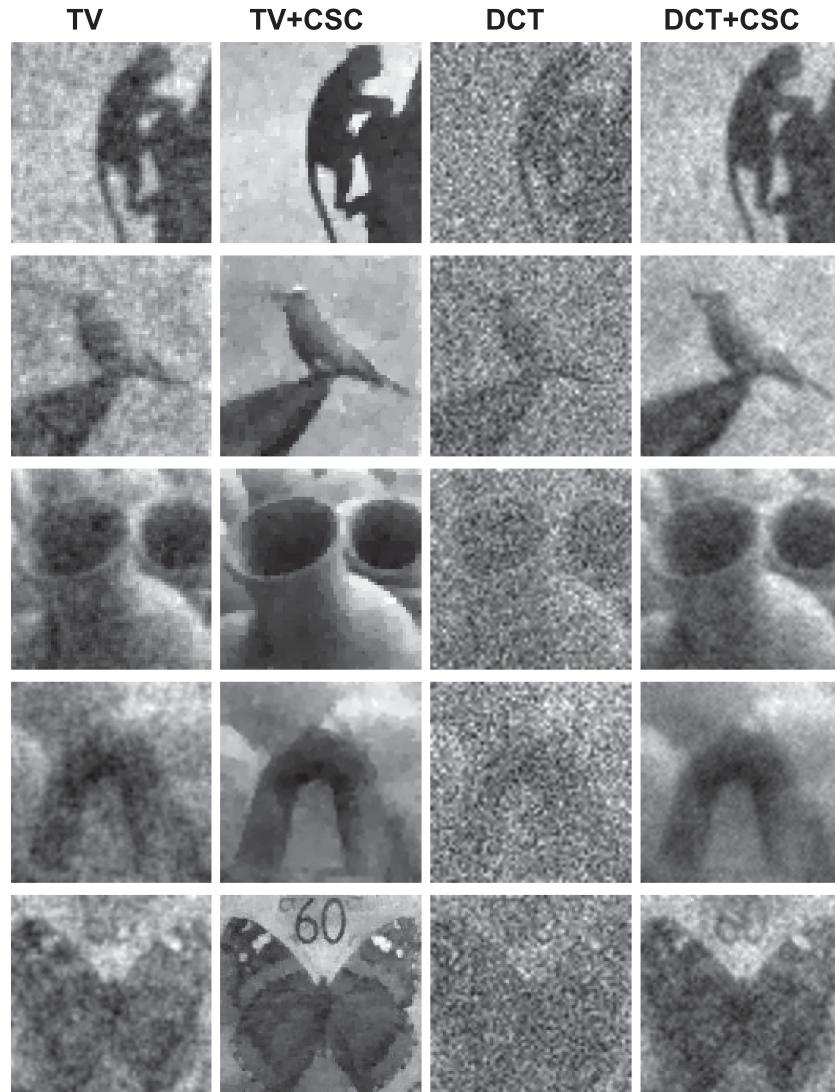


Fig. 6. The imaging results of five different scenes using our setup. The compression ratio is 0.25, and each measurement is averaged over 200 samplings for noise suppression.

when the SNR exceeds 70 dB. As aforementioned, since the DCT algorithms performs well for reconstructing periodic structures, while TV related methods show advantages over DCT related ones, especially under higher SNR. The promotion after CSC introduction is increasing as the noise level decreases, both for TV and DCT prior. This result would provide guidance in actual experiment: minimising system noise would achieve higher promotion through introducing CSC prior additionally.

4. Experiments on Real Captured Data

Finally, we experimentally demonstrate the performance of our method on a SPI prototype. The scheme of our experimental setup is shown in Fig. 5. The light emitted from the halogen lamp is first collimated by a beam collimator composed of a condense lens, an optical integrator and a shaping lens. After collimation, a digital micro-mirror device (DMD) modulates the incoming light with random patterns. The outgoing light from the DMD is expanded by the projecting lens as scene

illumination. After interaction with the target scene, the outgoing beam is converged by a collecting lens and finally captured by a bucket detector.

In implementation, the pixel resolution is 128×128 , consistent with the simulation. We set the compression ratio as 0.25 and capture approximately 4100 measurements. During the capture, several factors might influence the final reconstruction, such as the fluctuation of light source and background light, the instability of detector sensitivity, and minor vibration of the light path. To effectively suppress these system noise, we record each measurement by averaging over 200 samplings, and choose the maximum amplification gain of the bucket detector. As shown in Fig. 6, although there still exist some artifacts in the reconstruction, the reconstruction after introducing CSC prior demonstrates significant superiority over conventional methods. Benefiting from the introduced local prior, the reconstruction at 0.25 compression ratio is decent: the noisy background is suppressed and local details are reconstructed well. For example, the first four images exhibit sharper edges and cleaner background after CSC introduction; the number “60” in the image “Butterfly” can be obviously resolved in TV+CSC. In sum, through imposing sparse constraints under the SPI framework, the reconstruction quality could be improved significantly.

5. Conclusion

In conclusion, this paper proposes a high fidelity reconstruction algorithm for single-pixel imaging, by introducing convolutional sparse coding as a complementary counterpart to widely used global prior such as TV minimization and DCT coefficient sparsity. By incorporating above combinational constraints into the SPI framework, the reconstruction exhibits significant promotion and thus can decrease the required numbers of measurements in practical applications. The simulations tell that the average PSNR promotion at different compression ratios is approximately $2.5 \sim 6.4$ dB for TV and $3.4 \sim 13.7$ dB for DCT. Moreover, the experiments of our prototype also demonstrate the advantage of the proposed method.

In terms of computational load, the convolutional basis can be learned beforehand, and the time consuming convolutional operator in the optimization could be transformed to matrix multiplication in the Fourier domain. Therefore, the optimization in our algorithm would not introduce additional calculation complexity to conventional compressive sensing based methods.

Besides, our method is a general reconstruction algorithm for SPI, so it is compatible with most SPI setups. For example, we can introduce CSC into characterization of entangled photon pairs to improve the final reconstruction in quantum optics [33]. We can also learn the dictionary for a specific target scene for higher precision, and thus has the potential to broaden the practical application of single-pixel imaging.

References

- [1] T. Pittman, Y. Shih, D. Strekalov, and A. Sergienko, “Optical imaging by means of two-photon quantum entanglement,” *Phys. Rev. A*, vol. 52, pp. R3429–R3432, 1995.
- [2] R. S. Bennink, S. J. Bentley, and R. W. Boyd, “Two-photon coincidence imaging with a classical source,” *Phys. Rev. Lett.*, vol. 89, 2002, Art. no. 113601.
- [3] A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato, “Ghost imaging with thermal light: Comparing entanglement and classical correlation,” *Phys. Rev. Lett.*, vol. 93, 2004, Art. no. 093602.
- [4] A. Valencia, G. Scarcelli, M. D’Angelo, and Y. Shih, “Two-photon imaging with thermal light,” *Phys. Rev. Lett.*, vol. 94, 2005, Art. no. 063601.
- [5] N. Tian, Q. Guo, A. Wang, D. Xu, and L. Fu, “Fluorescence ghost imaging with pseudothermal light,” *Opt. Lett.*, vol. 36, pp. 3302–3304, 2011.
- [6] C. Zhao *et al.*, “Ghost imaging lidar via sparsity constraints,” *Appl. Phys. Lett.*, vol. 101, 2012, Art. no. 141123.
- [7] B. Sun *et al.*, “3D computational imaging with single-pixel detectors,” *Science*, vol. 340, pp. 844–847, 2013.
- [8] P. Clemente, V. Durán, E. Tajahuerce, and J. Lancis, “Optical encryption based on computational ghost imaging,” *Opt. Lett.*, vol. 35, pp. 2391–2393, 2010.
- [9] W. Chen and X. Chen, “Ghost imaging for three-dimensional optical security,” *Appl. Phys. Lett.*, vol. 103, 2013, Art. no. 221106.
- [10] O. S. Magana-Loaiza, G. A. Howland, M. Malik, J. C. Howell, and R. W. Boyd, “Compressive object tracking using entangled photons,” *Appl. Phys. Lett.*, vol. 102, 2013, Art. no. 231104.

- [11] 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.*, vol. 104, 2014, Art. no. 251120.
- [12] Y. Bromberg, O. Katz, and Y. Silberberg, "Ghost imaging with a single detector," *Phys. Rev. A* vol. 79, 2009, Art. no. 053840.
- [13] W. Gong and S. Han, "A method to improve the visibility of ghost images obtained by thermal light," *Phys. Lett. A*, vol. 374, pp. 1005–1008, 2010.
- [14] F. Ferri, D. Magatti, L. Lugiato, and A. Gatti, "Differential ghost imaging," *Phys. Rev. Lett.*, vol. 104, 2010, Art. no. 253603.
- [15] B. Sun, S. S. Welsh, M. P. Edgar, J. H. Shapiro, and M. J. Padgett, "Normalized ghost imaging," *Opt. Exp.*, vol. 20, pp. 16892–16901, 2012.
- [16] K. Guo, S. Jiang, and G. Zheng, "Multilayer fluorescence imaging on a single-pixel detector," *Biomed. Opt. Exp.*, vol. 7, pp. 2425–2431, 2016.
- [17] W. Wang, X. Hu, J. Liu, S. Zhang, J. Suo, and G. Situ, "Gerchberg–Saxton-like ghost imaging," *Opt. Exp.*, vol. 23, 2015, Art. no. 28416.
- [18] C. Deng, J. Suo, Y. Wang, Z. Zhang, and Q. Dai, "Single-shot thermal ghost imaging using wavelength-division multiplexing," *Appl. Phys. Lett.*, vol. 5, 2018, Art. no. 051107.
- [19] O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.*, vol. 95, 2009, Art. no. 131110.
- [20] M. Aßmann and M. Bayer, "Compressive adaptive computational ghost imaging," *Sci. Rep.*, vol. 3, 2013, Art. no. 1545.
- [21] L. Bian, J. Suo, Q. Dai, and F. Chen, "Experimental comparison of single-pixel imaging algorithms," *JOSA A*, vol. 35, pp. 78–87, 2018.
- [22] W. Yu, M. Li, X. Yao, X. Liu, L. Wu, and G. Zhai, "Adaptive compressive ghost imaging based on wavelet trees and sparse representation," *Opt. Exp.*, vol. 22, pp. 7133–7144, 2014.
- [23] W. Gong and S. Han, "High-resolution far-field ghost imaging via sparsity constraint," *Sci. Rep.*, vol. 5, 2015, Art. no. 9280.
- [24] J. Yang, J. Wright, T. Huang, and Y. M, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [25] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
- [26] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary learning for sparse coding: Algorithms and convergence analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1356–1369, Jul. 2016.
- [27] X. Hu, J. Suo, T. Yue, L. Bian, and Q. Dai, "Patch-primitive driven compressive ghost imaging," *Opt. Exp.*, vol. 23, pp. 11092–111042, 2015.
- [28] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Conf. Comput. Vis.*, 2015, pp. 1823–1831.
- [29] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 391–398.
- [30] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5135–5143.
- [31] 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.*, vol. 3, pp. 1–122, 2011.
- [32] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.
- [33] N. Montaut *et al.*, "Compressive characterization of telecom photon pairs in the spatial and spectral degrees of freedom," *Optica*, vol. 5, pp. 1418–1423, 2018.