Spatial-temporal compressive imaging using an unfolding network

Dingaoyu Zhao[1], Edmund Y. Lam[2], and Jun Ke[1,*]
[1]*Beijing Institute of Technology, School of Optics and Photonics, Beijing, China, 100081*
[2]*Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong SAR, China*
[*]*jke@bit.edu.cn*

Abstract: We propose an unfolding network GapUNet for spatial-temporal compressive imaging. Simulation and optical experiments demonstrate the network performance using compression ratios of 128 : 1 and 16 : 1. The mean PSNR of the reconstructed objects is higher than 29 dB.

1. Introduction

Existing compressive imaging approaches mainly focus on compression in one dimension, i.e., spatial [1] or temporal [5] domain. However, it is often desirable to use a low-speed and low-resolution camera to obtain high-quality images with better speed and spatial resolution. In this work, we study the reconstruction in spatial-temporal compressive imaging (STCI) [2] using an unfolding network GapUNet, which is designed based on an iterative algorithm GAP (Generalized Alternating Projection) [3]. Using the network, 8 frames of 1920×1080 images are reconstructed from one measurement frame of size 480×270 in 3 seconds. The average PSNR of the reconstructed objects from the DAVIS2017 testset is higher than 29 dB. We also demonstrate our idea with optical experiments.

2. Object reconstruction in STCI using GapUNet

The mathematical model for STCI can be written as

$$\mathsf{Y} = (\sum_{i}^{n_t} \mathsf{C}_i \odot \mathsf{X}_i)_{\downarrow s}, \tag{1}$$

where X_i and C_i are the ith object frame and mask of size $n_w \times n_h$, respectively. Y is the measurement frame of size $(\frac{n_w}{s} \times \frac{n_h}{s})$. The operation $\odot$ and $\downarrow_s$ represent the dot product and the downsampling process. The spatial and temporal compression ratios are $s^2 : 1$ and $n_t : 1$. Redefining the object and the measurement frames as vectors, we can have the measurement model as $\mathbf{y} = \mathsf{H}\mathbf{x}$.

The algorithm GAP has been studied for temporal compressive imaging (TCI), which has a similar measurement model as STCI. The optimization problem solved using GAP is defined as

$$(\hat{\mathbf{x}}, \hat{v}) = \underset{\mathbf{x},\mathbf{v}}{argmin} \frac{1}{2}\|\mathbf{x} - v\|_2^2 + \lambda g(v) \quad s.t. \quad \mathbf{y} = \mathsf{H}\mathbf{x}. \tag{2}$$

In this equation, $g(\cdot)$ is a penalty function. We solve it iteratively as follows, with k indicating the iteration number. Given v, $\mathbf{x}^{(k+1)}$ is updated via an Euclidean projection of v^k onto the linear manifold $\mathscr{M} : \mathbf{y} = \mathsf{H}\mathbf{x}$, such that

$$\mathbf{x}^{(k+1)} = v^k + \mathsf{H}^T(\mathsf{H}\mathsf{H}^T)^{-1}(\mathbf{y} - \mathsf{H}v^k). \tag{3}$$

Given $\mathbf{x}$, $v^{(k+1)}$ can be updated by a denoiser with $\delta = \sqrt{\lambda}$ representing the standard deviation of the additive white Gaussian noise,

$$v^{(k+1)} = \mathscr{D}_\delta(x^{(k+1)}). \tag{4}$$

To solve the reconstruction problem in STCI, we design two modules, $\mathscr{L}$ with fixed parameters and a trainable CNN-based denoiser $\mathscr{D}$, to replace Eq. 3 and Eq. 4. Similar to [4], we also design a module $\mathscr{H}$ to generate δ for each iteration and different compression ratios. $\mathscr{H}$ has three fully connected layers. The whole structure of the GapUNet is shown in Fig. 1(a).

3. Simulated and optical experiments

We train GapUNet using the dataset DAVIS2017, which consists of 90 image sequences in different scenes. These image sequences are cropped into $8\times120\times120$ patches without overlapping. We set parameters $n_t = 8$ and $s = 4$, which means less than 1% data are collected using STCI. Fig. 1(b) presents the reconstructions of three sequences. The sequence *Chamaleon* is rich in details. Objects in the sequence *Car-race* move fast. The last image sequence presents reconstruction PSNR over 30 dB.

We also conduct an optical experiment to verify our idea. A DMD DLP6500 is used to display the modulated high-speed and high-resolution object sequence. Then, a sensor Basler ACA720-520 is used to collect the measurements with a $4f$ optical system. The original objects have resolution 120×120. The measurement frame has size 60×60. The exposure time of the camera is set as 8ms. Four images are reconstructed from one measurement frame. Fig. 1(c) presents the results.

Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, pcAOP) @ Optica Publishing Group 2022

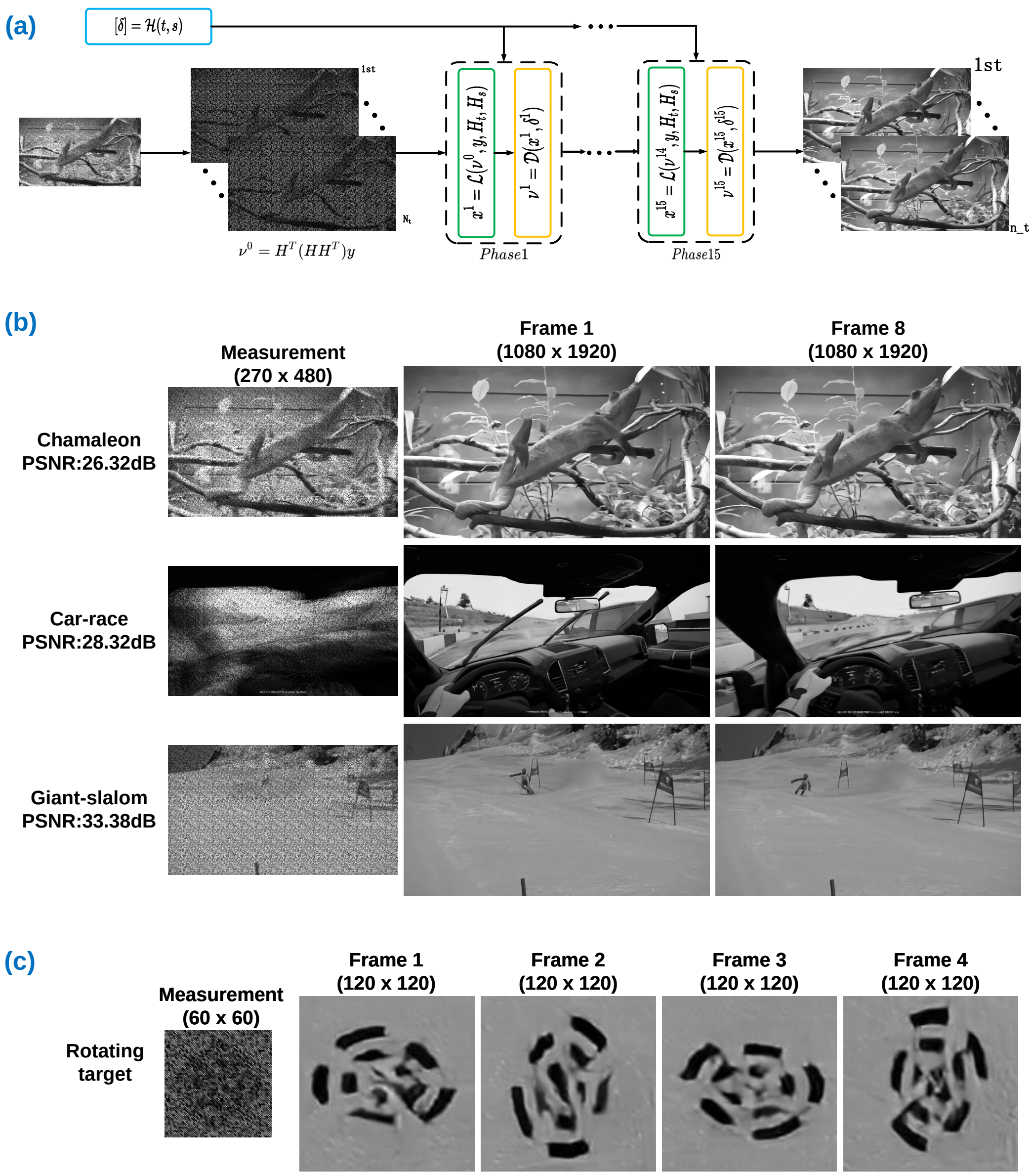

Fig. 1. (a)The structure of GapUNet with modules $\mathscr{H}$ (blue box), $\mathscr{L}$ (green box) and $\mathscr{D}$ (orange box); reconstructed results in (b) simulated and (c) optical experiments.

4. Conclusion

In summary, we design a network GapUNet for STCI to reconstruct high-resolution images with PSNR close to 30 dB when the compression ratio is 128. We also conduct an optical experiment to verify our idea when $n_t = 4$ and $s = 2$.

References

1. Can Cui and Jun Ke. Spatial compressive imaging deep learning framework using joint input of multi-frame measurements and degraded maps. *Opt. Express*, 30(2):1235–1248, Jan 2022.
2. Jun Ke, Linxia Zhang, and Edmund Y. Lam. Temporal compressed measurements for block-wise compressive imaging. In *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)*, page JW4B.1. Optica Publishing Group, 2019.
3. Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
4. Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
5. Linxia Zhang, Edmund Y. Lam, and Jun Ke. Temporal compressive imaging reconstruction based on a 3d-cnn network. *Opt. Express*, 30(3):3577–3591, Jan 2022.