applied optics

Compressive video sensing with side information

Xin Yuan,[1] Yangyang Sun,[2] and Shuo Pang[2,*]

[1]*Nokia Bell Labs, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA*
[2]*College of Optics and Photonics (CREOL), University of Central Florida, 4304 Scorpius St., Orlando, Florida 32816, USA*
Corresponding author: pang@creol.ucf.edu

Our temporally compressive imaging system reconstructs a high-speed image sequence from a single, coded snapshot. The reconstruction quality, similar to that of other compressive sensing systems, often depends on the structure of the measurement, as well as the choice of regularization. In this paper, we report a compressive video system that also captures the side information to aid in the reconstruction of high-speed scenes. The integration of the side information not only improves the quality of reconstruction, but also reduces the dependence of the reconstruction on regularization. We have implemented a system prototype that splits the field of view of a single camera into two channels: one channel captures the coded, low-frame-rate measurement for high-speed video reconstruction, and the other channel captures a direct measurement without coding as the side information. A joint reconstruction model is developed to recover the high-speed videos from the two channels. By analyzing both the experimental and the simulation results, the reconstructions with side information have demonstrated superior performances in terms of both the peak signal-to-noise ratio and structural similarity.   © 2017 Optical Society of America

OCIS codes: (100.0100) Image processing; (150.0150) Machine vision.

https://doi.org/10.1364/AO.56.002697

1. INTRODUCTION

Inspired by theoretical developments in compressive sensing [1], several compressive imaging systems in the time domain have been conceived and implemented, aiming to extract high-frame-rate videos from low-speed image sequences [2–13]. In these systems, to recover the high-frame-rate video, the measurements are encoded by a flutter shutter [2,3], a physical mask [4–6], or a digital mirror array [7–10] whose pattern varies at a rate faster than the frame rate of the camera. This technique has been extended to microscopy imaging [11–13]. Video compressive sensing based on the spatial-multiplexing camera has also been investigated [14,15]. Various reconstruction algorithms [16–23] have been proposed to solve the inversion problem, and the accuracy of the reconstruction relies heavily on the coded measurement. As the reconstruction extracts both the spatial and temporal information from a single measurement, the quality of the reconstruction would be compromised if the incorrect regularization is adopted. Towards the goal of improving the reconstruction quality and reducing the dependency on regularization, in this work, we designed a compressive temporal imaging system that includes side information (SI) [24–27] in the measurement: in addition to the temporal-spatial coded measurement, we simultaneously capture another measurement without coding by the same camera. The SI is a blurred version of a fast scene, i.e., a temporal summation of the high-speed video frames.

SI-aided compressive sensing has recently been investigated in both theory [24,25] and applications [26,27]. In this paper, we propose to use the simple blurry image as the SI for video compressive sensing. The setup is simple, and both cameras used in the system could be off-the-shelf low-speed cameras [4,5]. This SI-aided system improves the reconstruction results significantly while maintaining the advantages of video compressive sensing cameras [4–6].

This paper makes the following contributions: (i) a video compressive imaging system aided by capturing a low-speed image sequence as the SI. Specifically, we have implemented a system prototype that captures two measurements, one coded measurement and the other without any coding, by a single camera at a low frame rate. (ii) A joint video reconstruction framework is developed, where the low-speed image is used as the SI to enhance the quality of the reconstructed video frames. We compared the reconstruction with some well-known reconstruction algorithms at different compression ratios, and our results demonstrated superior performances in terms of both the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [28].

2. IMAGING SYSTEM AND FORWARD MODEL

Figure 1 depicts the schematic of the SI-aided imaging concept. Two measurements are captured for the same high-speed scene.

Fig. 1. Schematic of the compressive video sensing with SI.

The first measurement is captured using the coded aperture in the imaging space. The high-speed video frames are encoded with different shifting variants of the same physical mask [4,5]. On the other hand, we use the same camera to capture a low-speed *uncoded* image, which could be blurry due to the motion of the scene. Our system would use the captured SI to improve the reconstruction quality. It is worth noting that this SI can also work with other video compressive sensing setups, for example, using a digital mirror array [7–10] for spatial encoding.

A. Forward Model

The high-speed scene can be modeled as a three-dimensional function $f(x, y, t)$. Let $h^{(1)}(x, y, t)$ denote the mask imposed on the scene on the first sensing channel at time t. Assuming the integration time of the camera is T, the first measurement can be expressed as

$$g^{(1)}(x, y) = \int_0^T f(x, y, t) h^{(1)}(x, y, t) \mathrm{d}t. \qquad (1)$$

As the second measurement is not modulated during the integration time, we have

$$g^{(2)}(x, y) = \int_0^T f(x, y, t) \mathrm{d}t. \qquad (2)$$

Since the mask is moving continuously, each high-speed frame is modulated by a different pattern [4,5]. The forward model is depicted in Fig. 2.
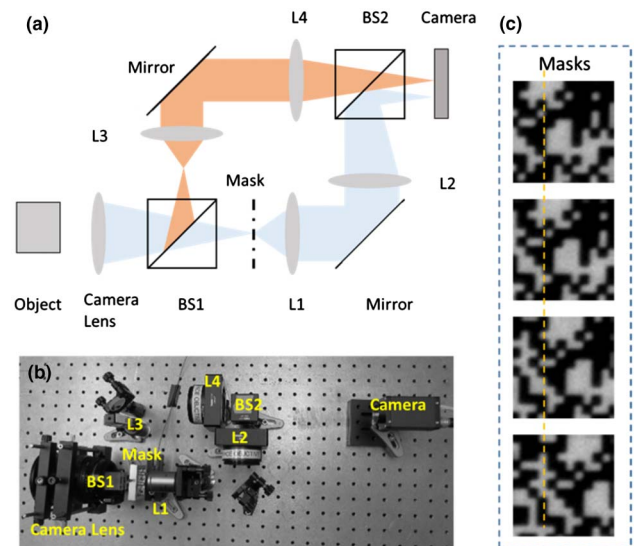
B. System Setup

The optical system mismatches between the channels and camera synchronization are the main challenges for the system implementation. We designed a system that uses the same detector to capture both the coded measurement channel and the SI channel. The system setup is shown in Fig. 3(a).

The system places a beam splitter (BS1) to separate the same scene into two channels. The camera lens (Nikon, 18–35 mm) is used as the objective lens for both channels. The intermediate
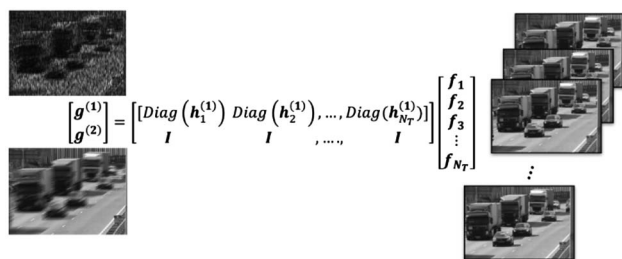
Fig. 2. Forward model of the imaging system. The left side shows two measurements (coded image and blurry image), and the fast scene is shown on the right.

Fig. 3. System setup. (a) The schematic of the prototype: the intermediate image by the camera lens is split by a beam splitter (BS1) into two intermediate images. One path is modulated by a high-speed mask and relayed to the camera; the other path is directly imaged to the camera. Two paths are combined by another beam splitter (BS2). (b) A photo of the prototype. (c) The mask at 4 locations, recorded on the same area of the detector.

image in the main channel is modulated by a binary random mask (HTA, Photomask) driven by a piezo-actuator (PI, P-840). The relay system in the main path consists of two lenses: L1 (Nikon, $4 \times f = 50$ mm) and L2 (Newport, $f = 200$ mm). In the SI path, L3 (Newport, $f = 50.2$ mm) and L4 (Newport, $f = 200$ mm) are used to match the optics used in the main path. The images from the two paths are merged on the same detector (AVT, G-145B) through the second beam splitter (BS2). A slight tilt of the mirror in the SI path separates the measurements. The separation of the two channels is approximately 550 pixels on the camera. The f-numbers of the two paths are both 5. The camera and the piezo-actuator are synchronized and triggered by a digital acquisition board (NI, USB6353). A 5 mm × 6 mm piece of black cardboard with a white letter "U" serving as the object is translated at 5 mm/s by a motor-driven actuator (Newport, LTA—HS). Figure 3(b) shows a photo of the system.

It is worth noting that our design shown in Fig. 3 is an example of our idea of using SI to boost the video reconstruction quality. In this design, we split the CCD into two parts in order to capture the main measurement and SI simultaneously. This is achieved by the extra hardware (mirrors and lenses) and field-of-view sacrifice. Other designs of SI and hardware implementations are discussed in Section 3.B.

To construct the forward model, we measure the transmission function of the mask by placing it on a white calibration background. The images of the mask at 4 different steps are shown in Fig. 3(c).

3. RECONSTRUCTION

We discretize the high-speed frames f, masks $h^{(1)}$, and measurements $g^{(1)}$, $g^{(2)}$. Considering that there are N_T high-speed

frames inside the integration time T, we denote the kth, $k = 1, ..., N_T$ discretized frame as $F_k \in R^{n_x \times n_y}$, which has $n_x \times n_y$ pixels. Similarly, $H_k^{(1)}$ is the kth mask pattern. For each pixel (i, j), $i = 1, ..., n_x; j = 1, ..., n_y$, we have

$$g_{i,j}^{(1)} = \sum_{k=1}^{N_T} h_{i,j,k}^{(1)} f_{i,j,k}, \tag{3}$$

$$g_{i,j}^{(2)} = \sum_{k=1}^{N_T} f_{i,j,k}. \tag{4}$$

Let f_k be the vectorized form of F_k. The vectorized form of the first measurement can be expressed as

$$g^{(1)} = H^{(1)} f, \tag{5}$$

$$f = [f_1 \quad f_2 \quad \cdots \quad f_3]^T, \tag{6}$$

where $H^{(1)}$ is the sensing matrix of the first measurement. Let $h_k^{(1)}$ denote the kth vectorized mask, and $H^{(1)}$ can be expressed as

$$H^{(1)} = [\mathrm{Diag}(h_1^{(1)}) \mathrm{Diag}(h_2^{(1)}), ..., \mathrm{Diag}(h_{N_T}^{(1)})], \tag{7}$$

where $\mathrm{Diag}(h_k^{(1)})$ denotes a diagonal matrix whose diagonal elements are composed of the vector $h_k^{(1)}$. Similarly,

$$g^{(2)} = [I, I, ..., I] f, \tag{8}$$

where I denotes the identity matrix with dimensions $(n_x n_y) \times (n_x n_y)$. We define $H^{(2)} = [I, I, ..., I]$, and concatenate both measurements as follows:

$$g = Hf, \tag{9}$$

$$g = \begin{bmatrix} g^{(1)} \\ g^{(2)} \end{bmatrix}, \qquad H = \begin{bmatrix} H^{(1)} \\ \alpha H^{(2)} \end{bmatrix}, \tag{10}$$

where α is a constant to scale the measurements to the same illumination level.

The reconstruction problem can be formulated as

$$\hat{f} = \arg_f \min \| g - Hf \|_2^2 + \tau\, R(f). \tag{11}$$

$R(f)$ denotes the regularizer, and it can be used to impose the sparsity of the signal on the basis of wavelet and the discrete cosine transformation (DCT) [4,5] or the total variation (TV) operator [11,17,22]. The regularizer penalizes reconstructions that do not conform to the characteristics of the estimated f. τ is the Lagrange parameter that balances the measurement error [the first term in Eq. (11)] and the regularizer.

Without the SI $g^{(2)}$, the problem reduces to the typical temporal compressive video problem [4,5],

$$\hat{f} = \arg_f \min \| g^{(1)} - H^{(1)} f \|_2^2 + \tau R(f). \tag{12}$$

A. Total Variation Regularization

Wavelet plus DCT priors have demonstrated excellent performance in video compressive sensing inversion problems [4,5,29], which usually need specific spatial sizes of the video for the algorithms to be efficient. The TV-based algorithms [11,17,22,30] do not have this limitation and can be used in any spatial size of video. Furthermore, the generalized alternating projection (GAP) [16] -based TV algorithm has

demonstrated a superior performance to other algorithms also using TV regularizers [17], especially in video and hyperspectral compressive sensing inversion problems. When the TV method is used, the regularizer in Eq. (10) will be

$$R(f) = \sum_{k=1}^{N_T} \sum_{i,j}^{n_x, n_y} \sqrt{(f_{i+1,j,k} - f_{i,j,k})^2 + (f_{i,j+1,k} - f_{i,j,k})^2}. \tag{13}$$

Let D denote the pixel-wise differentiation operator, and $z = Df$. The developed GAP-TV is an iteration algorithm. For tth-iteration, introducing the intermediate variables w, v, the updated equations for GAP-TV consist of the following steps:

$$w_{t+1} = f_t + \mu H^T (HH^T)^{-1} (g - Hf_t), \tag{14}$$

$$v_{t+1} = w_{t+1} - D^T z_t, \tag{15}$$

$$f_{t+1} = \mathrm{clip}\left(z_t + \frac{1}{\beta} Dv_{t+1}, \frac{\gamma}{2} \right), \tag{16}$$

where the clipping function is defined as

$$\mathrm{clip}(b, P) := \begin{cases} b & |b| \le P \\ P\, \mathrm{sign}(b) & |b| > P \end{cases}, \tag{17}$$

and β, γ are the parameters used in the clipping function. $\mu \in (0, 2]$ is the step size [31]. After initializing $z_0 = 0$, $f_0 = 0$, we run Eqs. (14)–(16) iteratively until the termination criteria are satisfied.

It is worth noting that in Eq. (13), we need to compute the inverse of HH^T. Fortunately, both $H^{(2)} H^{(2)T}$ and $H^{(1)} H^{(1)T}$ are diagonal matrices:

$$H^{(1)} H^{(1)T} = \mathrm{Diag}[c_1^2, c_2^2, ..., c_{n_x n_y}^2], \tag{18}$$

$$H^{(2)} H^{(2)T} = N_T I_{n_x n_y}. \tag{19}$$

We have

$$(HH^T)^{-1} = \begin{bmatrix} H^{(1)} H^{(1)T} & \alpha H^{(1)} H^{(2)T} \\ \alpha H^{(2)} H^{(1)T} & \alpha^2 H^{(2)} H^{(2)T} \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}, \tag{20}$$

where

$B_1 = (H^{(1)} H^{(1)T} - H^{(1)} H^{(2)T} (H^{(2)} H^{(2)T})^{-1} H^{(2)} H^{(1)T})^{-1}$,

$B_4 = (\alpha^2 H^{(2)} H^{(2)T}$
$\quad -\alpha^2 H^{(2)} H^{(1)T} (H^{(1)} H^{(1)T})^{-1} H^{(1)} H^{(2)T})^{-1}$,

$B_2 = -\alpha (H^{(1)} H^{(1)T})^{-1} H^{(1)} H^{(2)T} B_4$,

$B_3 = -\alpha^{-1} (H^{(2)} H^{(2)T})^{-1} H^{(2)} H^{(1)T} B_1$.

As $\{B_1, B_2, B_3, B_4\}$ are all diagonal matrices and can be precomputed and stored, the inverse of HH^T can be obtained in closed form. Therefore, the GAP-TV algorithm is efficient for our applications of video compressive sensing with SI, and it is solely composed of Eqs. (14)–(16). As Eqs. (15) and (16) can be recognized as the TV-denoising approach, other algorithms, such as IST [32] and TwIST [30], can also be used to update

w in Eq. (14), leading to IST-TV and TwIST-TV. Other TV-based denoising algorithms can be used [30] as well.

B. OTHER TYPES OF SIDE INFORMATION

We have developed a prototype and algorithms for video compressive sensing with SI. We have divided the CCD into two sections in our design shown in Fig. 3, where we aim to utilize a single CCD to capture both the main measurement and the side information. However, it is not necessary to do so, as we can easily replace the two mirrors in Fig. 3 with two cameras. Though extra effort is required for synchronization between these two cameras, it will mitigate the large size problem of the CCD and the cost of the focal plane array (FPA) beyond-visible bandwidth.

While the proposed SI is a blurred image of the high-speed scene, other types of SI can also be used. We further consider the following two cases: (1) the SI is an independent coded measurement of the main channel, i.e., two independent coding patterns are imposed on the 2 channels at the same time, and (2) the SI is a coded image but coded by a flutter shutter [2]. The first case can be considered that we have 2 independent measurements for the same high-speed scene. As the SI channel also needs a mechanism to encode the scene, the cost and power of the system will be increased. In the second case, the SI is the summarization of the high-speed scene when the shutter is open; therefore, it is partial information of our proposed SI in Section 2. However, this regime needs the shutter of the SI channel to be operated faster than the camera, and this would increase the cost of the entire imaging system as well. We verify the feasibility of these two kinds of SI and compare the performance with the proposed one via a simulation in Section 4.

From the algorithmic perspective, the only change using these two different kinds of SI in the forward model is $H^{(2)}$ in Eq. (10). As $H^{(2)}H^{(2)T}$ is still a diagonal matrix, a slight change is required to perform a reconstruction using these two types of SI.

C. COMPARISON WITH OTHER RECONSTRUCTION ALGORITHMS

The reconstruction algorithms proposed above do not consider using other training data that have been used in the dictionary learning [7,23,33], GMM [18,19,34,35], and deep learning [23] -based reconstruction algorithms. These learning-based algorithms have achieved excellent reconstruction results. However, the quality of the reconstructed videos relies on the training data. This paper aims to show that we can get improved reconstruction results with low-cost SI. The TV-based algorithms developed above only impose piecewise smooth priors to the video frames. The following simulation demonstrates the enhanced performance introduced by the SI. We have also checked that other algorithms, especially the GMM-based algorithms, will also be improved by using SI. Since a comparison of various reconstruction algorithms and priors is out of the scope of this paper, in the following experimental results, we only report the results using TV-based reconstruction algorithms. However, the observation is ready to be generalized into other reconstruction regimes.

4. SIMULATION RESULTS

In this section, we conduct simulation to verify the hardware principle and the proposed reconstruction algorithm. We randomly generate the binary mask {0, 1} with equal probability [18,19]. High-speed videos are synthesized or downloaded from some datasets as used in Ref. [18]. Different compression rates N_T are used to generate the measurements $\{g^{(1)}, g^{(2)}\}$, where $g^{(1)}$ is the summation of the coded high-speed video frames and $g^{(2)}$ is just the summation of these frames. With these measurements as input, we run the algorithms to get the reconstructed video frames. The PSNR and SSIM are employed as metrics to evaluate the quality of reconstructed videos. The SSIM measures the similarity between images and is designed to improve on traditional metrics, such as PSNR and mean squared error (MSE) which have shown to be inconsistent with human visual perception [28]. In all the reconstruction simulations, the regularization parameter was adjusted to provide the optimal results.

We first test the proposed imaging system and reconstruction algorithms with a synthesized video, as shown in Fig. 4. Each video frame is composed of two parts: the top part is a rotating fan with "UCF," and the bottom part is a still panel with "BELL." The video frame is of spatial size 192×192. The two generated measurements are also shown in Fig. 4. Reconstructed video frames using different algorithms are shown on the right side of Fig. 4, and the PSNR and SSIM are summarized in Fig. 5. We test both $N_T = 4$ and $N_T = 10$. It can be observed from Fig. 5 that for all three algorithms, the reconstructed videos with SI have higher PSNRs (about 2 dB) and SSIMs (about 0.02) than the reconstructed videos without SI, which is in agreement with our analysis. Furthermore, we notice that the GAP-TV algorithm with SI always performs best in terms of both PSNR and SSIM. In particular, the GAP-TV with SI has a PSNR that is improved by ~7 dB when $N_T = 4$ and ~3.75 dB when $N_T = 10$.

Next, we perform the simulations on the videos used in [18], namely, the "Traffic," "Train," and "Dunk" videos. We use $N_T = 8$ for all three of these videos. The spatial sizes of "Traffic" and "Train" are 256×256, while the "Dunk" video has a spatial size of 512×512. The PSNRs and SSIMs of the reconstructed videos are summarized in Fig. 6. Similar to the "BELL-UCF" video, the PSNR and SSIM have been improved using SI for all the algorithms. In particular, TwIST-TV has been improved more than 9 dB in terms of the PSNR for

Fig. 4. Measurements and reconstructed video frames with different algorithms for synthesized "BELL-UCF" video with $N_T = 4$ compared with the ground truth. Notice the quality improvement with SI inside the red rectangles.
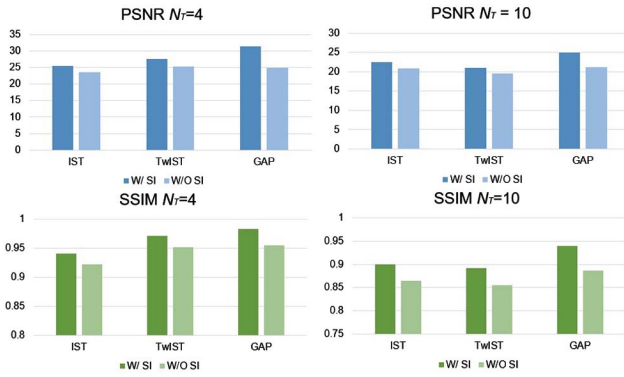
Fig. 5. PSNR and SSIM of reconstructed video frames with different algorithms for the synthesized "BELL-UCF" video with $N_T = \{4, 10\}$.
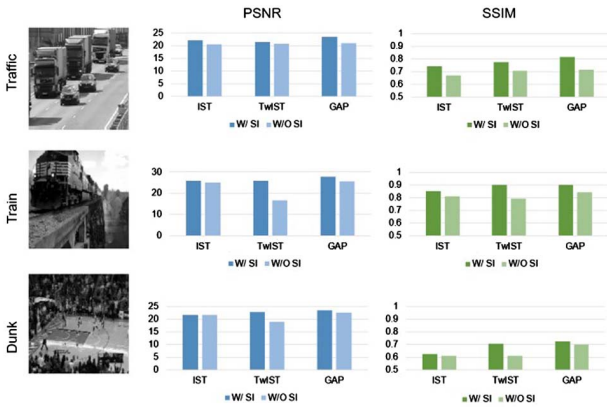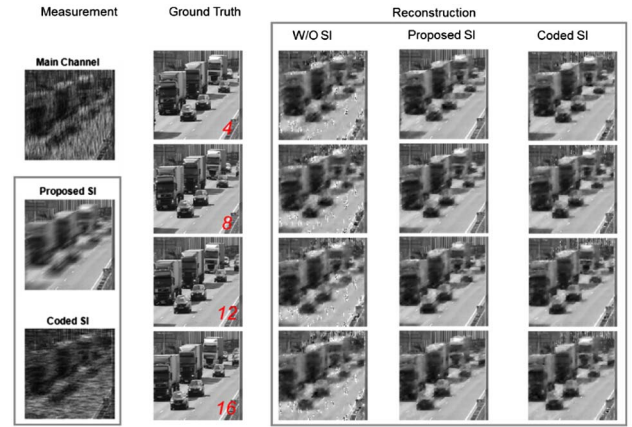

Fig. 6. Reconstructed PSNR and SSIM of "Traffic," "Train," and "Dunk" video for $N_T = 8$.

the "Train" video. Example reconstructed video frames of the "Traffic" video are shown in Fig. 7. Again, GAP-TV with SI provides the best reconstruction, i.e., providing details as well as the precise motions of the cars. Similar cases exist in the other videos. Notice the vertical shapes (red rectangles in Fig. 7) inside the video frames; all the reconstructed frames with SI provide more details than the results without SI.

We now compare the performance of our proposed SI with the other types of SI described in Section 3.B, namely the coded SI, where the second channel is independently modulated.
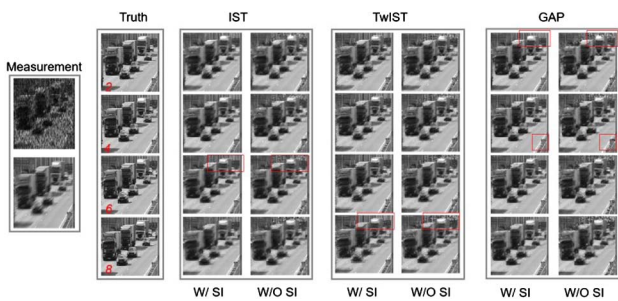
Fig. 7. Measurements and reconstructed frames with various algorithms for "Traffic" video with $N_T = 8$.

Fig. 8. Selected reconstructed video frames of the "Traffic" video with $N_T = 16$ using different types of SI. Frames 4, 8, 12, and 16 are plotted for visualization. The PSNR and SSIM are shown in Fig. 9.
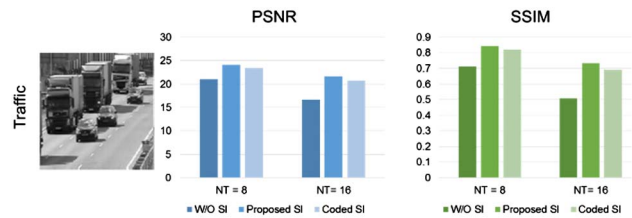

Fig. 9. PSNR and SSIM of reconstructed "Traffic" video with $N_T = 8, 16$ using the proposed SI and the coded SI.

Since the flutter-shutter SI is solely partial information of our proposed SI, it consistently produced worse reconstruction results. We thus only show the results of proposed SI and coded SI in Figs. 8 and 9, where we use the "Traffic" video as an example, and GAP-TV is employed to perform the reconstruction. It can be seen that, though the coded SI (and flutter-shutter SI) are more complicated than the proposed SI, their results are not as good as those of our proposed SI. We suspect that the proposed SI provides complementary information to the main measurement, which is designed to capture the motion, while the proposed SI is essential for still images.

We further investigate how the SI will help the reconstruction in a high compression case, for example, $N_T = 16$ in Fig. 8, where we demonstrated the results of the "Traffic" video with the proposed SI and coded SI. It can be seen that without SI, the reconstructed video frames have severe artifacts. By contrast, when the SI is utilized in the reconstruction, the reconstruction quality improved significantly. Specifically, the proposed SI increases the PSNR by 7 dB (Fig. 8 and the far right part of Fig. 9), 1 dB better than the coded SI. This again confirmed that complementarity is important for SI. The coded SI, though increasing the measurement number, does not provide more information about the still image and therefore would not perform as well as our proposed SI for the entire video.

5. EXPERIMENTAL RESULTS

We applied the imaging system prototype described in Section 2.B to capture a high-speed scene and validate our
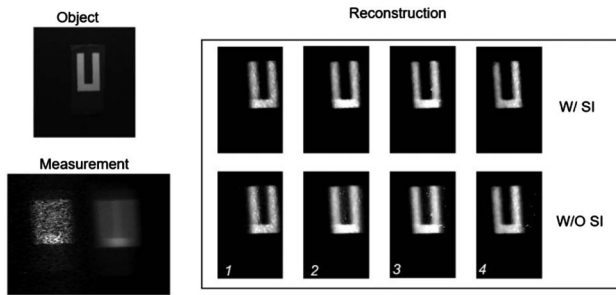
Fig. 10. Experimental scene (top left), measurement (bottom left), and reconstructed frames (right) with $N_T = 4$. It can be seen the SI improves the results significantly, especially at the edges of the letter.
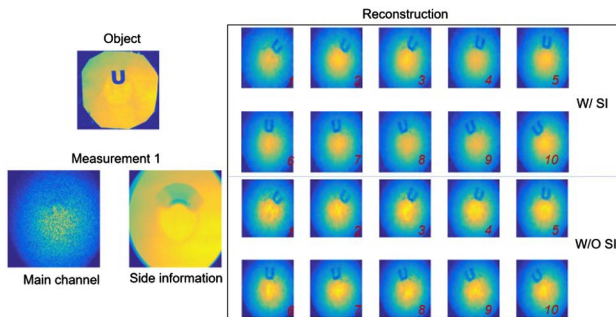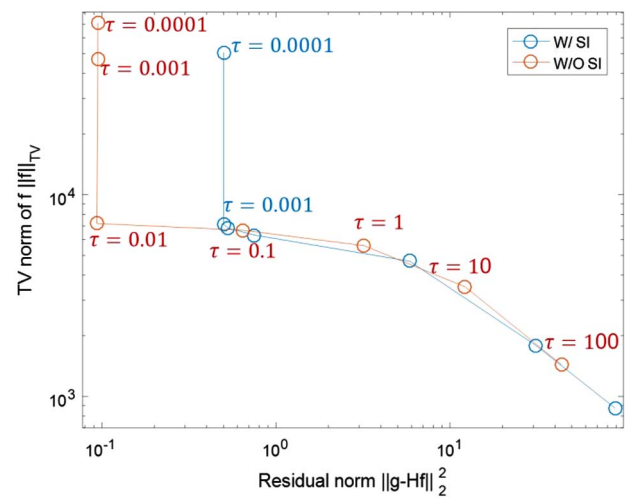

Fig. 11. Experimental object (top left), measurements (bottom left), and reconstructed frames (right) with a compressed rate of $N_T = 10$. The reconstruction with SI (top right) is compared with the reconstruction without SI (bottom right).

Fig. 12. *L*-curve of the "BELL-UCF" video with and without using SI.

reconstruction algorithm. We have reconstructed 4 frames from a single snapshot of both the coded measurement and the SI. The scene is a "*U*" on a panel (top left in Fig. 10) shifting from left to right. The measurement captured by our camera is shown in the bottom left part of Fig. 10. The reconstruction results with and without SI are shown in the right part of Fig. 10. Since the ground truth is not available for this real video, we evaluate it by visual perception. It can be observed that the reconstruction with SI is improved, with fewer visible speckle artifacts.

To further verify the performance of the SI at high-speed motion, we built another hardware using a digital mirror device [7,8] to encode the main channel. We capture a high-speed scene in Fig. 11, where a "*U*" is on a panel spinning in the anti-clockwise direction (top left in Fig. 11). It can be seen from the measurements that the scene is blurred, and we cannot identify the motion. From the reconstructed 10 frames in the right part of Fig. 11, we can clearly identify the motion. Furthermore, the SI information has improved the results significantly, again, on the edge of the letter and the panel. The camera is working at 60 fps and, thus, the reconstructed video is at a speed of 600 fps.

6. DISCUSSION

Both the experiments and simulation showed superior reconstruction results with the SI. Even though the SI does

not have a high resolution in the temporal domain, it amounts to adding extra constraints in the spatial domain, especially for the intensity, which was not directly available from the sparse measurement. We use a graphical tool, the so-called *L*-curve, to further analyze the relation between the SI and the regularization. The *L*-curve is a plot of the regularization norm, $\|\mathbf{f}\|_{\mathrm{TV}}$, of the regularized solution versus the corresponding residual norm $\|\mathbf{g}-H\mathbf{f}\|_2$ for all valid Lagrange coefficients, as shown in Fig. 12. It displays the compromise between the minimization of the regularization error and the reconstruction error, which is at the heart of any regularization method [36]. When very little regularization is introduced, the error is dominated by the perturbation error, which corresponds to the uppermost part of the *L*-curve, above the middle corner. When a large amount of regularization is introduced, most filter factors are small, and the error is dominated by the regularization error, which corresponds to the rightmost part of the *L*-curve. In between, there is a region where both contribute comparably, and this region defines the *L*-shaped corner. We plot the *L*-curve for the situations with and without SI, with Lagrange coefficient varying from 10^{-4} to 10^2. From Fig. 11, we can observe that at the "under-smoothed" region (above the middle corner of the L-curve), the residual norm of the reconstruction with SI is greater than the one without SI, because the reconstruction with SI has two measurement channels. It is worth noting that the location of the corner of the *L*-curve is often used as an approximation to the optimal regularization parameter. The optimized Lagrange coefficient is in the order of 10^{-3} with SI, compared with 10^{-2} for the single coded measurement. This indicates that with the SI, the dependence of the reconstruction on the regularization is reduced. More analyses can be found in [25].

In the aspect of system implementation, although using a single aperture can minimize the difference between two channels, small differences still exist. The magnification of the two relay paths becomes obvious at the pixel level. The magnification difference indicates the sensing matrix of the blurry path is not an identity matrix. In our experiment, we calibrate the

magnification difference with a checkerboard as the object. The numerical aperture of the lens and the aberrations of the two channels also have slight differences. In the setup, considering the expense of the system, we used optics with a lower resolution in the main path so the aberrations of the mask would be minimized.

7. CONCLUSION

We have reported a video compressive sensing system that captures high-speed videos at a low frame rate. Two measurements are captured simultaneously in one snapshot: one measurement is a coded compressive measurement, and the other is a simple low-speed image that is the summation of the high-speed scene. The second measurement is easy to obtain, serving as the SI in our system to aid reconstruction. A joint reconstruction algorithm is developed to recover high-speed videos from these measurements. The simulation and experimental results have demonstrated improved performances of the reconstructed video with the SI.

Recently, we also investigated the scenario where the two measurements are measuring the scene from different perspectives, which could provide the range information [37–40] and polarization [41] of the scene and has potential in high-speed stereo imaging applications [42]. A similar idea can also be used in other bandwidth imaging systems, e.g., x rays [43–46].

REFERENCES

1. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).
2. J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe, "Flutter shutter video camera for compressive sensing of videos," in *IEEE International Conference on Computational Photography (ICCP)*, Seattle, Washington, 2012, pp. 1–9.
3. A. Veeraraghavan, D. Reddy, and R. Raskar, "Coded strobing photography: Compressive sensing of high speed periodic videos," IEEE Trans. Pattern Anal. Mach. Intell. **33**, 671–686 (2011).
4. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," Opt. Express **21**, 10526–10706 (2013).
5. X. Yuan, P. Llull, X. Liao, J. Yang, G. Sapiro, D. J. Brady, and L. Carin, "Low-cost compressive sensing for color video and depth," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, 2014, pp. 3318–3325.
6. T.-H. Tsai, P. Llull, X. Yuan, D. J. Brady, and L. Carin, "Spectral-temporal compressive imaging," Opt. Lett. **40**, 4054–4057 (2015).
7. Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *IEEE Conference on Computer Vision and Pattern Recognition*, Barcelona, Spain, 2011, pp. 287–294.
8. D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *IEEE Conference on Computer Vision and Pattern Recognition*, Denver, Colorado, 2011, pp. 329–336.
9. R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. Katsaggelos, "High spatio-temporal resolution video with compressed sensing," Opt. Express **23**, 15992–16007 (2015).
10. L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**, 74–77 (2014).
11. X. Yuan and S. Pang, "Compressive video microscope via structured illumination," in *IEEE International Conference on Image Processing (ICIP)* (2016), pp. 1589–1593.
12. A. Stevens, L. Kovarik, P. Abellan, X. Yuan, L. Carin, and N. D. Browning, "TEM video compressive sensing," Microsc. Microanal. **21**, 1583–1584 (2015).
13. X. Yuan and S. Pang, "Structured illumination temporal compressive microscopy," Biomed. Opt. Express **7**, 746–758 (2016).
14. A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk, "CS-MUVI: Video compressive sensing for spatial-multiplexing cameras," in *IEEE International Conference on Computational Photography* (2012), pp. 1–10.
15. T. Goldstein, L. Xu, K. F. Kelly, and R. Baraniuk, "The STOne transform: Multi-resolution image enhancement and compressive video," IEEE Trans. Image Process. **24**, 5581–5593 (2015).
16. X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted-minimization with applications to model-based compressive sensing," SIAM J. Imaging Sci. **7**, 797–823 (2014).
17. X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *IEEE International Conference on Image Processing (ICIP)*, Phoenix, Arizona, 2016, pp. 2539–2543.
18. J. Yang, X. Yuan, X. Liao, P. Llull, G. Sapiro, D. J. Brady, and L. Carin, "Video compressive sensing using Gaussian mixture models," IEEE Trans. Image Process. **23**, 4863–4878 (2014).
19. J. Yang, X. Liao, X. Yuan, P. Llull, G. Sapiro, D. J. Brady, and L. Carin, "Compressive sensing by learning a Gaussian mixture model from measurements," IEEE Trans. Image Process. **24**, 106–119 (2015).
20. X. Yuan, J. Yang, P. Llull, X. Liao, G. Sapiro, D. J. Brady, and L. Carin, "Adaptive temporal compressive sensing for video," in *International Conference on Image Processing (ICIP)* (2013), pp. 14–18.
21. P. Llull, X. Yuan, X. Liao, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Temporal compressive sensing for video," in *Compressed Sensing and Its Applications: MATHEON Workshop 2013* (Springer 2015), pp. 41–74.
22. H. Schaeffer, Y. Yang, and S. Osher, "Space-time regularization for video decompression," SIAM J. Imaging Sci. **8**, 373–402 (2015).
23. M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," arXiv:1603.04930 (2016).
24. F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," IEEE Trans. Inf. Theory **62**, 6459–6492 (2016).
25. J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, Georgia, 2014, pp. 512–516.
26. X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. J. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," IEEE J. Sel. Top. Signal Process. **9**, 964–976 (2015).
27. G. A. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar, "Adaptive-rate compressive sensing using side information," IEEE Trans. Image Process. **24**, 3846–3857 (2015).
28. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
29. X. Yuan, P. Llull, D. Brady, and L. Carin, "Tree-structure Bayesian compressive sensing for video," arXiv:1410.3080 (2014).
30. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).
31. X. Yuan, H. Jiang, G. Huang, and P. Wilford, "SLOPE: Shrinkage of local overlapping patches estimator for lensless compressive imaging," IEEE Sens. J. **16**, 8091–8102 (2016).
32. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**, 183–202 (2009).

33. X. Yuan, "Compressive dynamic range imaging via Bayesian shrinkage dictionary learning," Opt. Eng. **55**, 123110 (2016).

34. J. Yang, X. Yuan, X. Liao, P. Llull, G. Sapiro, D. J. Brady, and L. Carin, "Gaussian mixture models for video compressive sensing," in *IEEE International Conference on Image Processing* (2013), pp. 15–18.

35. X. Yuan, H. Jiang, G. Huang, and P. A. Wilford, "Compressive sensing via low-rank Gaussian mixture models," arXiv:1508.06901 (2015).

36. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," SIAM Rev. **34**, 561–580 (1992).

37. P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Image translation for single-shot focal tomography," Optica **2**, 822–825 (2015).

38. X. Yuan, X. Liao, P. Llull, D. Brady, and L. Carin, "Efficient patch-based approach for compressive depth imaging," Appl. Opt. **55**, 7556–7564 (2016).

39. Y. Sun, X. Yuan, and S. Pang, "High-speed compressive range imaging based on active illumination," Opt. Express **24**, 22836–22846 (2016).

40. P. Llull, X. Yuan, X. Liao, J. Yang, L. Carin, G. Sapiro, and D. Brady, "Compressive extended depth of field using image space coding," in *Computational Optical Sensing and Imaging (COSI)* (2014), pp. 1–3.

41. T.-H. Tsai, X. Yuan, and D. J. Brady, "Spatial light modulator based color polarization imaging," Opt. Express **23**, 11912–11926 (2015).

42. X. Yuan, Y. Sun, and S. Pang, "Compressive temporal stereo-vision imaging," in *Imaging and Applied Optics*, OSA Technical Digest (Optical Society of America, 2016), paper JT3A.30.

43. S. Pang, M. Hassan, J. Greenberg, A. Holmgren, K. Krishnamurthy, and D. Brady, "Complementary coded apertures for 4-dimensional x-ray coherent scatter imaging," Opt. Express **22**, 22925–22936 (2014).

44. L. Wang, J. Huang, X. Yuan, K. Krishnamurthy, J. Greenberg, V. Cevher, D. J. Brady, M. Rodrigues, R. Calderbank, and L. Carin, "Signal recovery and system calibration from multiple compressive Poisson measurements," SIAM J. Imaging Sci. **8**, 1923–1954 (2015).

45. J. Huang, X. Yuan, and R. Calderbank, "Collaborative compressive x-ray image reconstruction," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, April, 2015, pp. 3282–3286.

46. J. Huang, X. Yuan, and R. Calderbank, "Multi-scale Bayesian reconstruction of compressive x-ray image," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, April, 2015, pp. 1618–1622.