Research Article | Vol. 29, No. 9 / 26 April 2021 / Optics Express 12787
Optics EXPRESS

Compressive temporal imaging using a rolling shutter camera array

FELIPE GUZMÁN,[1] PABLO MEZA,[2] AND ESTEBAN VERA[1,*]

[1] School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[2] Department of Electrical Engineering, Universidad de La Frontera, Temuco, Chile
[*] esteban.vera@pucv.cl

Abstract: In this work, we present a novel camera array that exploits the electronic rolling shutter to achieve high speed compressive temporal imaging. Traditional compressive temporal imaging makes use of mechanical coded apertures, despite implementation and calibration challenges. Instead, we propose to model the inherent spatial and temporal coding provided by the distinctive rolling shutter sampling from each camera of the array as a compressive temporal imaging system matrix. Thus, we can recover a high speed video from a set of snapshots from the camera array by using compressive sensing reconstruction algorithms. We present both simulation and experimental results for a 4-camera array system with different orientation angles, reconstructing up to 56 high-speed sub-frames from a set of simultaneously triggered snapshots from the array, achieving a compression rate of up to 14X.

1. Introduction

For most video applications, a temporal resolution of 30fps is often enough since human vision is not able to distinguish fast changes or movements beyond that. However, there are many applications where a higher temporal resolution analysis is highly desired for better understanding of natural and artificial processes [1–3]. Over the last forty years, the CCD sensor [4] has been widely used in the vast majority of imaging applications and all sorts of research, while the CMOS sensor [5] has been mostly adopted as a low-cost, low-quality alternative for consumer cameras. Nonetheless, recent improvements in semiconductor fabrication processes, plus a variety of intrinsic advantages offered by the CMOS sensor technology [6], has turned it into a viable and popular alternative not only for consumer imaging, but also for highly demanding scientific imaging [7].

One of the main issues found in CMOS sensors is related to its readout circuitry–mostly due to the time-multiplexing needed for reading every pixel row–named as Rolling Shutter (RS). In essence, the RS reads different rows of the detector at different moments within the global integration time of the detector, creating a variety of unwanted distortions when imaging dynamic events such as skew, wobble, and aliasing, among others. These distortions arise from either global or local movements such as fast camera movements or fast moving objects, respectively. Even though the RS can be seen as a nuisance to the imaging pipeline, some research has taken advantage of it to increase the overall temporal resolution by synchronizing a massive CMOS camera array [8], delivering a temporal resolution of up to 1560fps. Another alternative proposed in [9] was the direct modification to the readout circuitry in the CMOS sensor to swap sampling rows and also define a different exposure time per row, allowing to apparently increase the temporal resolution by better sampling the space-time datacube. However, as simple as it may sound, it is not yet an off-the-shelf CMOS detector solution.

On the other hand, another alternative to increase the temporal sampling of the imaging system is through the use of compressed sensing [10]. For example, Llull *et al.* [11] proposed a modified imaging system by adding a moving coded aperture that encodes the high-speed scene within the integration time of the detector. By properly calibrating the space-time sampling of the coded

#418892 https://doi.org/10.1364/OE.418892
Journal © 2021 Received 6 Jan 2021; revised 3 Mar 2021; accepted 2 Apr 2021; published 12 Apr 2021

Optics EXPRESS

aperture, several video frames can be reconstructed from a single snapshot using compressed sensing reconstruction techniques. Also, Pournaghi *et al.* [12] introduced a compressive camera-array system with various sampling methods to overcome the temporal limitations of the system. With advances in compressive video sensing, Zhang *et al.* [13] proposed a pixel-wise code exposure sensor for temporal imaging, which can output 100fps videos from 5fps measurements. Lately, another attempt for compressive temporal imaging has been proposed in [14], revisiting the opportunities brought by the RS while using a lensless computational imaging system based on a diffuser that acts as a pseudorandom mask, spatially spreading the scene such that the information of the entire scene is encoded in each row, dramatically increasing the temporal resolution for simple video scenes.

In this work, we present a novel camera array that exploits the electronic rolling shutter to achieve high speed compressive temporal imaging. Instead of using traditional compressive temporal imaging techniques that makes use of mechanical coded apertures, nor making any modification to the optics, we propose to exploit the inherent spatial and temporal coding provided by the rolling shutter itself at distinctive camera orientations, improving the sampling of the space-time datacube. Our contributions are the development of a forward matrix modeling of the RS acquisition process to simulate distortions in dynamic scenes and solve the compressive reconstruction problem posed by the proposed novel compressive temporal imaging system that exploits the built-in electronics of cheap off-the-shelf CMOS cameras, without the need of optical modifications neither coded apertures.

2. Rolling shutter modeling

No matter what class of array detector is being used, the acquisition process for a single detector row is mostly dominated by three temporal factors: the reset time, the exposure time, and the readout time. As shown in Fig. 1, if the three steps are simultaneous, then we are in presence of a Global (electronic) Shutter (GS). On the other hand, during the CMOS Rolling Shutter acquisition the timing for each row is displaced to assure that only one row is being read at a given time.

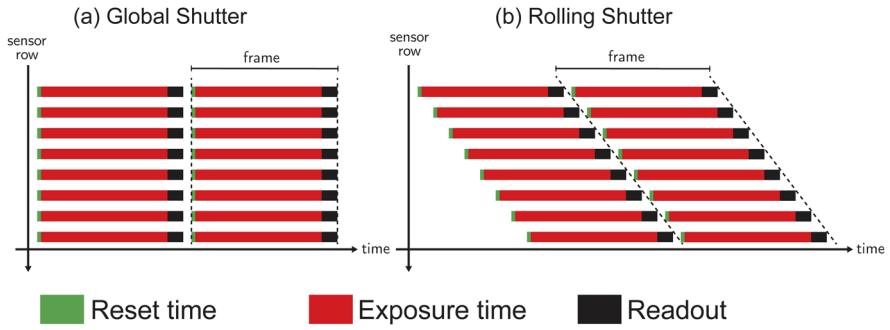

Fig. 1. Comparative representation of the acquisition system in detector arrays. (a) Global shutter; (b) Rolling shutter.

The acquisition system for an imaging detector can be modeled as a temporal integral at the sensor plane such that

$$P(x,y) = \int_0^\infty V(x,y,t) \cdot S(x,y,t)dt, \tag{1}$$

where (x, y) are the discrete pixel indices of the sensor, $V(x, y, t)$ is the time-varying optical intensity distributed on the image sensor plane (i.e. space-time datacube), $S(x, y, t) \in \{0, 1\}$ represents the binary shutter function. The characteristics of $S(x, y, t)$ are determined by the

Optics EXPRESS

acquisition parameters of the RS sensor: The readout time t_r, which is determined by the reading circuit clock speed. The exposure time t_e, which most of the sensors on the market allow to control. The delay time t_d, which is the time between each frame. Finally, the pixel size of the sensor $(\Delta r_x, \Delta r_y)$, as well as the detector resolution–or selected Region-of-Interest (ROI)–also plays a role.

Typically, the reset time is rather short (around $\sim 0.7\mu$s [15]) and, for simplicity, it is neglected. Also an isometric pixel geometry is assumed ($\Delta r_x \equiv \Delta r_y$). As shown in Fig. 2, a continuous dynamic scene can be quantized in time to generate sub-frame values as determined by Δt. Moreover, the value $N_s = \lceil t_e/t_r \rceil$ represent the number of rows exposed during Δt. With this in consideration, the frame of a RS CMOS detector can be seen as the collapsed space-time datacube in the temporal axis. For short exposure times (e.g. Figure 2(a)), the multiplexing between sub-frames of the scene is low, while for longer exposures there is a greater number of pixels sampled from more sub-frames.

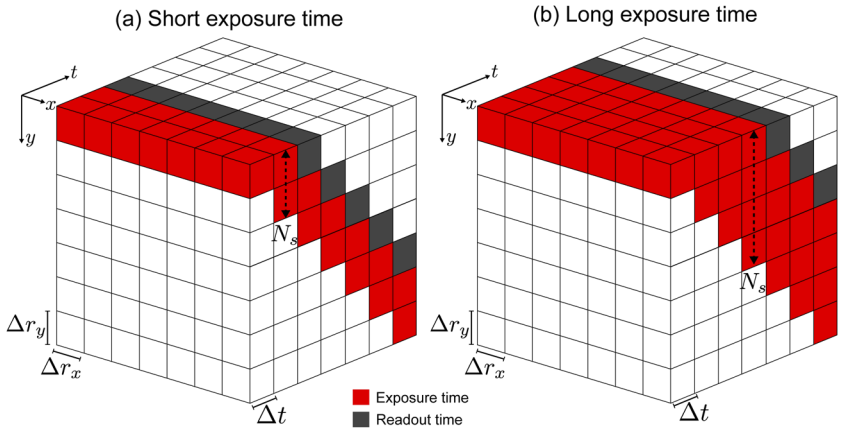

Fig. 2. Quantized representation of the Rolling Shutter acquisition. (a) Short exposure time ($t_e \simeq t_r$); (b) Long exposure time ($t_e > t_r$).

With the scheme presented in Fig. 2, it is easy to understand that when observing high-speed moving objects, these can be distorted depending on their direction relative to the readout of the sensor rows (scanline). Generally, for global movements (e.g. camera motion), distortions can be interpreted as purely geometric deformations. Nonetheless, for local movements different sorts of distortions can appear and may no longer be easily modeled.

In the discrete domain, we can reformulate Eq. (1) as the image formation of the $m \times n$ image I as follows

$$I(x, y) = \sum_{s=1}^{\mathrm{T}} S(x, y, s\Delta t, f) \odot V(x, y, s\Delta t, f), \tag{2}$$

where $\odot$ represents the Hadamard product and T all the sub-frames within the frame $f \in \{1, 2, \ldots, \mathrm{F}\}$ for F the total number of frames. In this approximation, all the temporal information in $[s\Delta t, (s+1)\Delta t]$ is integrated and not recoverable. For the model to work properly, $\Delta t \geq t_r$ since $\Delta t < t_r$ it will only add redundancy into the sub-frames. For simplicity, and for being the smaller value to be chosen without falling into redundancies, we use $\Delta t = t_r$.

In the context of RS, we first focus on a given row $y = k$ to define the temporal related variables as shown in Fig. 3, where we have the exposure time t_e and reset time t_r, plus the delay time variable t_d, which is the time between frames f. In a CMOS detector, the delay time can be chosen such that the exposure of the first row may start even before the readout of the last row of the previous frame. We will consider that the row readout time never overlaps each other,

Optics EXPRESS

so $t_d \geq mt_r - t_e$. The auxiliary variable t_α is the start acquisition time, which will be used to synchronize the proposed camera array in section 4.

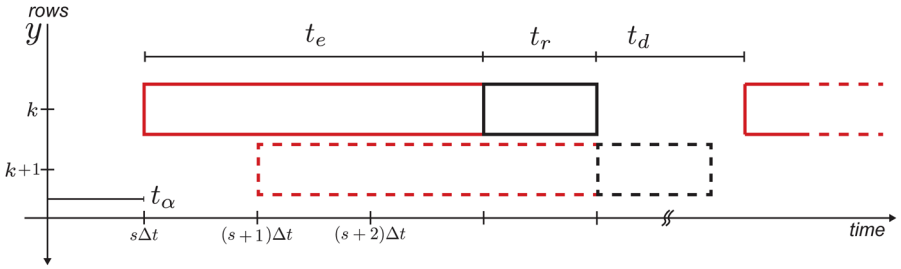

Fig. 3. Definition of the temporal parameters for the RS cycle.

Now, lets define the shutter function S as follows

$$S(x, y = k, s\Delta t, f) = \mathbb{M}_{s\Delta t}(x, y = k, f), \tag{3}$$

where $\mathbb{M}_{s\Delta t}$ is a binary mask with values defined by the conditions imposed by Eq. (4). This mask is the sampling matrix of the space-time datacube, and it is defined as

$$\mathbb{M}_{s\Delta t}(x, y = k, f) = \begin{cases} 1, & if \quad s\Delta t \in [t_\alpha,\ t_e + t_\alpha] + (k-1)t_r + t_d(f-1) \\ 0. & otherwise \end{cases} \tag{4}$$

With the proposed model, it is possible to simulate real conditions of high-speed scenes acquired with an RS sensor with great accuracy. For example, in Fig. 4(a) we have one frame of a rotating object captured with GS, while in Fig. 4(b) we have the same moving scene captured with RS. Using the GS image in Fig. 4(a) with the proposed shutter model together with the knowledge of the exact RS parameters used in the real RS snapshot of Fig. 4(b), we simulated what would be the resulting RS image as shown in Fig. 4(c), presenting great similarity to the originally captured RS image. By changing the high-speed input video V to the simulator, it is possible to generate all the distortions associated with the RS, and by controlling the exposure time we can make the simulation to naturally include motion blur, saturation and signal to noise.

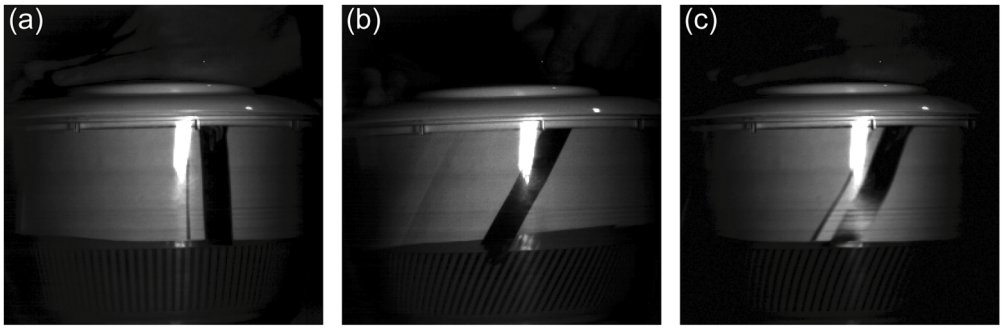

Fig. 4. Example of a black belt moving from right to left. (a) real GS snapshot; (b) real CMOS RS image at $t_e = 26\mu s$; (c) simulated RS image starting from (a) and using the same parameters of the RS detector in (b).

Optics EXPRESS

3. Compressive temporal imaging

Compressive temporal imaging enables high-speed imaging using normal speed detectors by modulating the space-time datacube with a dynamic coding aperture. As originally proposed in the CACTI system [11], the coding was implemented at a relayed image plane by a moving random aperture mask that acted over the full spatial span of the scene within the exposure time t_e. The forward model for the acquisition of a single-shot measurement is as follows

$$g(x,y) = \sum_{s=1}^{\mathrm{T}} S(x,y,s\Delta t) \odot V(x,y,s\Delta t) + n(x,y), \tag{5}$$

where n represent the noise of the system, and S is the coded aperture of the random shifted mask. Figure 5 shows a graphical representation of both the RS acquisition model and the CACTI. After comparing the image formation models in either Eqs. (2) and (5), it can be stated that they share an analog mathematical model, while having a different design for the sampling matrix S.

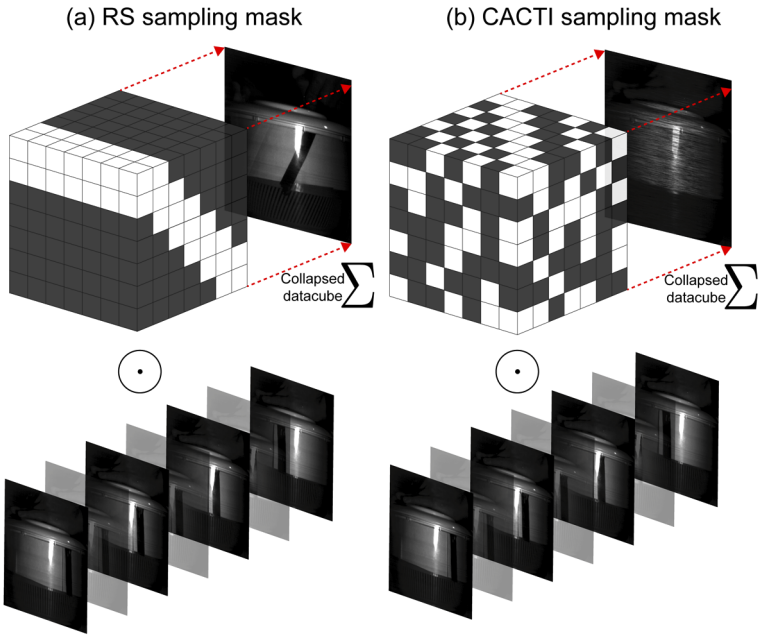

Fig. 5. Comparison of the RS (a) and CACTI (b) imaging systems. The final image I in each case correspond to the collapse in the time domain of the sampled datacube $S \odot V$. (a) S comes from the RS model; (b) S is a shifting spatial random mask.

We consider the spacetime datacube in lexicographical form as $\mathbf{f} \in \mathbb{R}^{mn\mathrm{T}\times 1}$, the measured image I as $\mathbf{g} \in \mathbb{R}^{mn\times 1}$ and the noise $\mathbf{n} \in \mathbb{R}^{mn\times 1}$. Now, the vectorized form of S becomes the forward matrix $\mathbf{H} \in \mathbb{R}^{mn\times mn\mathrm{T}}$, leading to the following linear forward model

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{6}$$

which represents the classic sensing model in compressed sensing [16]. Now, if the individual sampling matrices $\mathbf{h_s}$ for each coded frame are obtained by

$$\mathbf{h}_s \overset{\text{def}}{=} \operatorname{diag}\left([S(1,1,s\Delta t)\ S(2,1,s\Delta t)\cdots S(m,n,s\Delta t)]\right), \quad s = 1,\ldots,\mathrm{T}, \tag{7}$$

then, the forward matrix $\mathbf{H}$ is defined as follows

$$\mathbf{H}_i \overset{\text{def}}{=} [\mathbf{h}_1\mathbf{h}_2\cdots\mathbf{h}_{\mathrm{T}}]. \tag{8}$$

Optics EXPRESS

When comparing the forward matrices for the RS and CACTI imaging systems, we can notice that the sampling of the RS is uniform, allowing only a particular row to be sampled at a given sub-frame, while for the CACTI pixels are sampled from all over the image. As seen in Fig. 6(a), the number of pixels sampled by the RS correspond to $n \times N_s$. On the other hand, by using a random mask with transmittance of 50% (as in the example of Fig. 6(b)), the CACTI has $0.5 \cdot m \times n$ pixels sampled at every sub-frame.

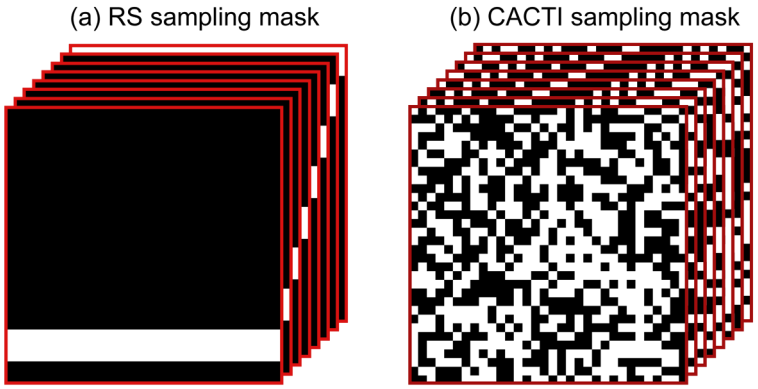

Fig. 6. Space-time sampling masks comparison. (a) RS system; (b) CACTI system.

To improve the sampling diversity provided by a single RS camera, we propose the use of an array of RS cameras at different orientations for compressive temporal image.

4. RS camera array

The fact that the RS provides with a mechanism to sample the space-time datacube, turns it into a potential candidate for compressive temporal imaging. Nevertheless, the main limitation of the RS acquisition system is the lack of diversity and the low number of pixels sampled at any particular sub-frame. To overcome this issue, we extend the single RS sampling (previously shown in Fig. 2) to handle an array of RS cameras oriented with different angles α_i relative to a reference camera.

In particular, we show in Fig. 7 the proposed scheme for an arrangement of four RS cameras placed at a distance z from the scene and spaced at a distance d between them, being the red camera the selected reference one. The angles α_i for each camera in the array are [0° 90° 180° 270°], respectively.

The sampling mask for the space-time datacube displayed in Fig. 7 is subject to the cameras having the same Field of View (FoV) and with a distance to the scene $z \gg d$ to avoid distortions due to changes in perspective determined by the adjacent but different positions of the cameras. Moreover, a tilt adjustment is applied so that the projections coincide at the common FoV. Consequently, extending the linear system in Eq. (6) to model the proposed camera array structure leads to the concatenated measurements

$$\underline{\mathbf{g}} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_C \end{bmatrix} = \underline{\mathbf{H}}\mathbf{f} + \mathbf{n}, \tag{9}$$

Optics EXPRESS

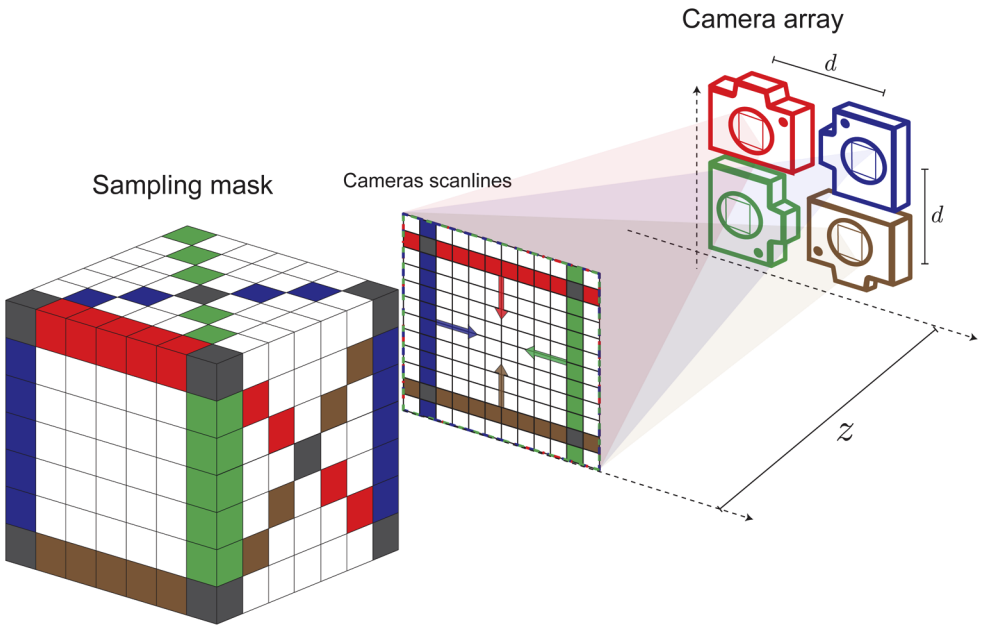

Fig. 7. Representation diagram of the proposed RS camera array displaying the relative position and dynamics of the scanlines within the space-time datacube. The sampling mask is represented by different colors (red, blue, brown and green) for each camera.

with the new forward matrix $\underline{\mathbf{H}}$ defined as follows

$$\underline{\mathbf{H}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_C \end{bmatrix}, \tag{10}$$

where $\mathbf{H}_i$ represent each individual forward matrix for the i^{th} camera. In particular, Eq. (10) requires the following conditions:

- The same optics is used in each camera so that the geomtric projections have the same characteristics.
- The same optical aperture is set in the lenses to keep the same photon flux, otherwise a scale factor must be applied to match the throughput of each $\mathbf{H}_i$
- Spatial calibration of each camera pose to adjust for the common ROI is performed to guarantee that each $\mathbf{H}_i$ is acquiring the same scene $\mathbf{f}$
- The acquisition is properly synchronized using controlled triggers for all cameras
- The characteristics of the detectors are similar in terms of readout noise, gain, and quantum efficiency [17–19]), and if not, they must be compensated by a proper detector calibration.

Optics EXPRESS

5. Results

5.1. Proposed reconstruction algorithm

Given the forward lineal model presented in Eq. (9) for the RS camera array, the reconstruction of the compressed space-time datacube is presented as the following unconstrained optimization problem

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\underline{\mathbf{g}} - \underline{\mathbf{H}}\mathbf{f}\|_2^2 + \lambda R(\mathbf{f}), \tag{11}$$

where R is the regularization function and λ the regularization weight parameter. In compressive temporal imaging, the regularization R is used to enforce sparsity on the reconstructed datacube $\mathbf{f}$ by using a representation basis for the spatial domain–such as Wavelet, DCT, or the Total Variation (TV) operator–where the redudancies are well studied. Even though there are several new algorithms for state-of-the-art compressive video reconstruction [20], the use of generalized alternating projection [21] (GAP), based on the alternating direction method of multipliers [22] (ADMM) algorithm, can achieve satisfactory results [20,23–25] with a reduced computational complexity. Therefore, the following results were rendered using the ADMM algorithm coupled with a TV regularization approach (ADMM-TV [26]), which can be described as the combination of the following optimization problems:

$$\mathbf{f}^{t+1} = \underset{\mathbf{f}}{\operatorname{argmin}} \frac{1}{2}\|\underline{\mathbf{g}} - \underline{\mathbf{H}}\mathbf{f}\|_2^2 + \frac{\eta}{2}\left\|\mathbf{f} - \left(\boldsymbol{\theta}^t + \mathbf{b}^t\right)\right\|_2^2, \tag{12}$$

$$\boldsymbol{\theta}^{t+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{\eta}{2}\left\|\boldsymbol{\theta} - \left(\mathbf{f}^{t+1} - \boldsymbol{b}^t\right)\right\|_2^2 + \lambda \mathrm{TV}(\boldsymbol{\theta}), \tag{13}$$

$$\boldsymbol{b}^{t+1} = \boldsymbol{b}^t - \left(\mathbf{f}^{t+1} - \boldsymbol{\theta}^{t+1}\right), \tag{14}$$

where $\mathbf{f}^{(t+1)}$ is updated as an euclidean projection of $\boldsymbol{\theta}^{(t)}$ as follows

$$\mathbf{f}^{(t+1)} = \left(\boldsymbol{\theta}^t + b^t\right) + \underline{\mathbf{H}}^\top \left[\frac{\underline{\mathbf{g}}_1 - \left[\underline{\mathbf{H}}\left(\boldsymbol{\theta}^t + b^t\right)\right]_1}{\eta + \psi_1}, \ldots, \frac{\underline{\mathbf{g}}_n - \left[\underline{\mathbf{H}}\left(\boldsymbol{\theta}^t + b^t\right)\right]_n}{\eta + \psi_n}\right]^\top, \tag{15}$$

which can be calculated using the i^{th} element-wise division $\underline{\mathbf{g}}_i - \left[\underline{\mathbf{H}}\left(\boldsymbol{\theta}^t + b^t\right)\right]_i / (\eta - \psi_i)$ where ψ can be derived as

$$\psi_i = \begin{bmatrix} \sum_{k=1}^{\mathrm{T}} \mathbf{h}_{k,1,i,i}^2 \\ \sum_{k=1}^{\mathrm{T}} \mathbf{h}_{k,2,i,i}^2 \\ \vdots \\ \sum_{k=1}^{\mathrm{T}} \mathbf{h}_{k,C,i,i}^2 \end{bmatrix}. \tag{16}$$

Finally, the TV operator is used to solve $\mathbf{f}^{(t)} - \boldsymbol{b}^{(t)}$ through the Chambolle's minimization algorithm [27].

5.2. Simulation results

By taking into account the linear model presented in Eq. (9) and the ADMM-TV reconstruction algorithm, we performed simulations of the acquisition and reconstruction processes for a variety of short image sequences, such as the Milk dataset, which is a slow motion video of an object falling into a glass of milk, and the Car dataset, which is a video of a toy car crashing against a wall, to name a few. In order to reduce the memory load and better match the experimental setup, the resolution of each datacube was fixed to small resolution images of (50×50) at 80

Optics EXPRESS

frames, ensuring a good sampling density of the datacube in space and time. We used both the peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) metrics for the evaluation of the image quality of the reconstructed videos with respect to the ground truth video. Figure 8 shows the average reconstruction results when using a single RS camera and the proposed RS camera array using from two up to eight cameras. Each camera added to the array is included with a rotation angle with respect to the reference camera following the order $[0°, 90°, 180°, 270°, 45°, 135°, 225°, 315°]$.

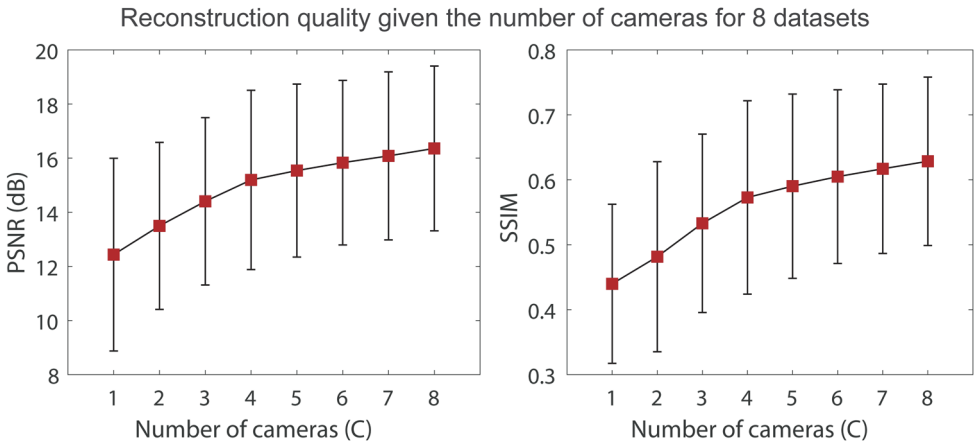

Fig. 8. Compressive reconstruction quality in terms of average PSNR and SSIM given the number of cameras in the RS camera array observing eight different moving scenes.

When considering only a single camera as a compressive temporal camera, the method struggles to provide with good results, which is expected due to the poor sampling. However, increasing the number of cameras helps to considerably increase the quality of the reconstructions. Overall, using more cameras is better, although the greatest impact is found when adding the first four cameras. Note that using more cameras would decrease the effective compression ratio.

Since the experimental results will be performed using a 4-camera array, we present several samples from the reconstructed videos for the simulated system in Fig. 9. With a mixture of simple and more complex movements, it can be seen that–for some scenes–the reconstruction is better when there is more sampling density or the speed of the scanline is comparable to the speed of the motion. Nonetheless, all reconstructed videos are able to show some degree of motion within the scene, which is not always easy to identify by solely observing the acquired static frames. After inspecting the video, the best result is achieved in the Milk dataset with a PSNR = 18.56dB for the sub-frame 48, moment when the object falls into the liquid, which can be fully appreciated. Another highlight is in the Traffic dataset, where the details of the trucks without any of the distortions associated with the RS can be observed, presenting a natural correction for the RS nuisance.

In addition, we present in Fig. 10 an analysis of the reconstruction of the Milk dataset when compared with 3D interpolation in both the temporal and spatial domains. Despite the reconstruction being far from perfect given the limited sampling of the rolling shutter camera array, we can notice that results are considerably better when using an sparse reconstruction algorithm in contrast with plain interpolation, demonstrating the compression ability of our proposed sampling scheme.

Optics EXPRESS

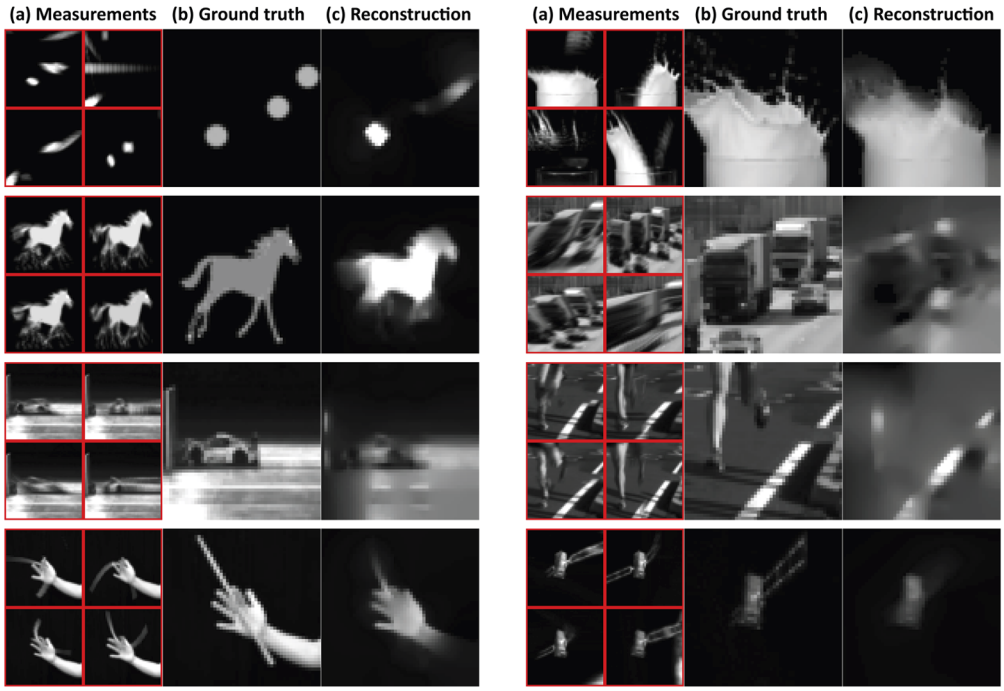

Fig. 9. Simulation results for eight datasets (see Visualization 1). (a) Simulation of the acquired snapshots with a 4-camera array with $t_e = 0.18$ ms, $t_r = 35\mu s$, generating $N_s = 5$. (b) Sample of the real video frames. (c) Sample of the reconstructed video frames with $\lambda = 20$, $\eta = 30$ and 200 iterations.

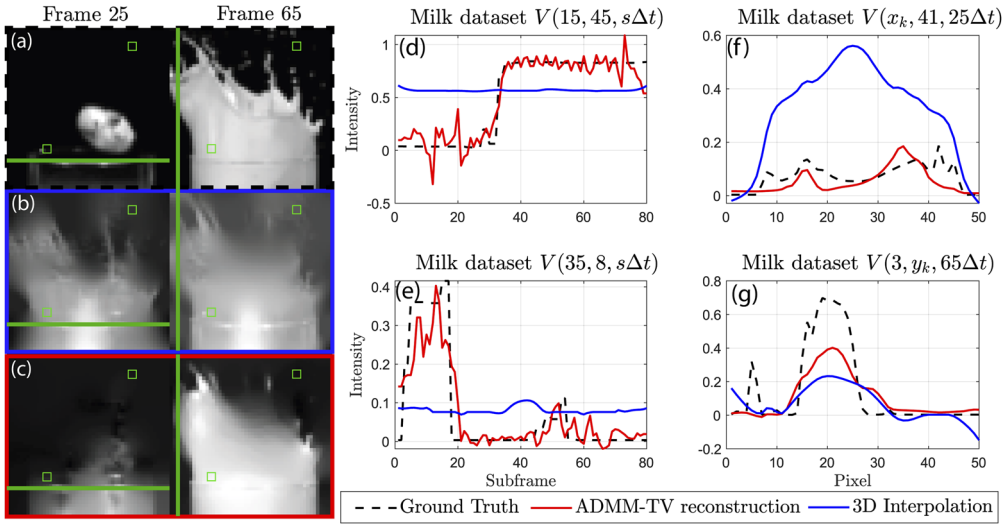

Fig. 10. Reconstruction analysis for the Milk dataset. Frame 25 and 65 (a) ground truth, (b) reconstructed with 3D interpolation and (c) reconstructed with ADMM-TV; (d) Temporal evolution of pixel (15,45); (e) Temporal evolution of pixel (35,8); (f) Spatial horizontal line of frame 25; (i) Spatial vertical line of frame 65.

Optics EXPRESS

5.3. Experimental results

To validate the proposed system for compressive temporal imaging, we built a 4-camera array test bed. The proof-of-concept imaging system is composed of four Basler Aca1300-30gm CMOS sensors with hybrid shutter, which have the option of selecting either global or rolling shutter as required, mounted with standard C-mount CCTV 25mm lenses, capturing images at 67 frames per second. The cameras were rotated following the proposed scheme in Fig. 7. The target scene consists of the light beam from an analog oscilloscope, which produces a relatively continuous movement without the need to worry about the screen refresh rate as in digital oscilloscopes, allowing to control the speed of the beam and its trajectory. The ROI calibration is performed with the help of four black points marked on the oscilloscope screen at the corners of the scene to be recorded. The camera setup also allows for a small tilt of the cameras which helps in putting the scene in the center of each camera, thus diminishing optical distortion differences. Then, ROIs are selected for each camera in order to have the marks imaged on the corner pixels.

As a first experiment, we present the results when using only two of the four available cameras, using as a scene a straight line beam movement with a slight inclination. One camera was oriented 90 degrees with respect to the other, selecting a common ROI from the scene of 40×40. The exposure time was set to $t_e = 200\mu$s and the readout time was experimentally measured to be 26μs, generating a value N_s of 7. From the two acquired snapshots shown in Fig. 11(a), we reconstructed 51 video frames using ADMM-TV as shown in Fig. 11(b). After visualizing the companion video, we can see that the beam is properly reconstructed crossing the scene in the right direction.

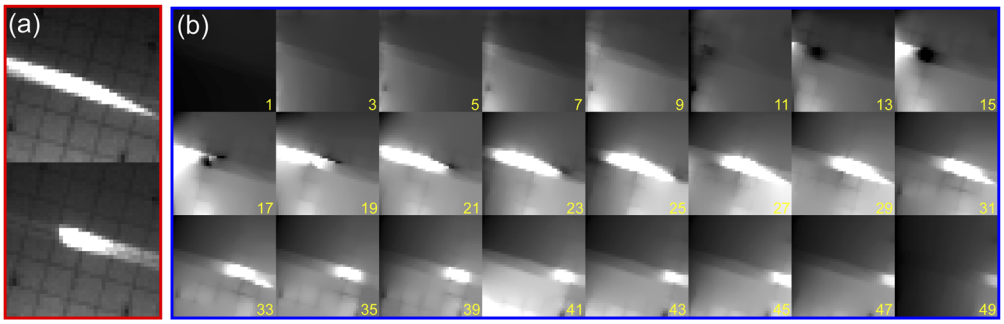

Fig. 11. Experimental results of a simple moving scene captured with a two-camera array (see Visualization 2). (a) RS captured images with the camera oriented with angles of 0° and 90°, respectively; (b) 24 out of 51 reconstructed frames with $\lambda = 20$, $\eta = 20$ and 450 iterations.

Finally, we acquired sample data using the 4-camera array system using identical temporal parameters as the last experiment. Two dynamic scenes using a triangular and a sinusoidal signal generated in the oscilloscope were captured. The oscilloscope was at a distance of 3 meters to minimize the differences between projections on the sensors and the time sweep of the oscilloscope beam was fixed at 20ms/DIV. With these parameters, the tangential speed of the beam is approximately 0.5m/s, compared to a calculated scanline speed of $\sim$ 0.2m/s. The acquired snapshots are shown in Fig. 12(a) and 12(c), while 21 of the 58 ADMM-TV reconstructed video frames (now of 50×50 pixels) are shown on Fig. 12(b) and 12(d).

After visualizing the video, we can appreciate that both beams trajectory and speed are well reconstructed. In particular, the triangular beam exhibit a better reconstruction than the sinusoidal beam, despite the fact that the beam changes dramatically in direction. Note that none of the presented results would actually surpass what can be achieved by state-of-the-art compressive temporal imaging devices. Nonetheless, we intended to demonstrate the great

Optics EXPRESS

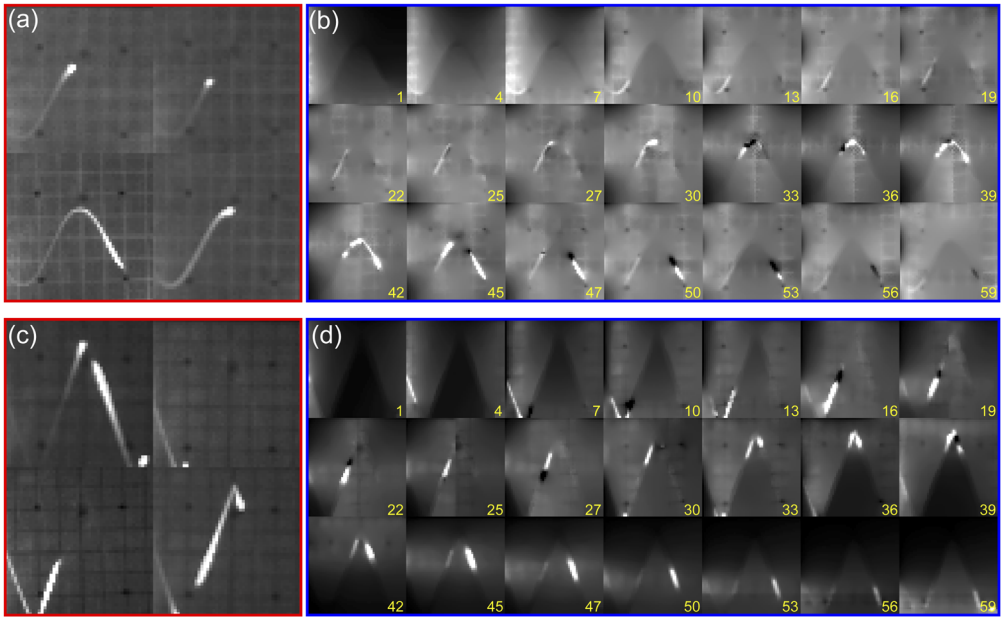

Fig. 12. Experimental results for two different scenes using the 4-camera array (see Visualization 3). (top) Sine wave scene measurements (a) and reconstructed frames (b) with $\lambda = 30$, $\eta = 20$ and 300 iterations; (bottom) Triangular wave scene measurements (c) and reconstructed frames (d) with $\lambda = 50$, $\eta = 22$ and 400 iterations.

advantage posed by using the RS as an electronic coding element at the hardware level, without needing optical/mechanical coding elements that often pose tough calibration challenges and a higher cost. Instead, for the proposed RS camera array, only a synchronized snapshot of the cameras and the previous characterization and knowledge of the RS parameters are needed to increase the temporal resolution, which could be of great usefulness for simple but fast motion within a scene.

Moreover, from the experimental results we can realize that the reconstruction quality is benefited when the velocity of the scene is comparable to the speed of the RS, which at a first glance limit the system ability, but it also opens the possibility of thinking in redesigning and better thinking the CMOS sensor to allow the control of the row/column addressing, somehow inspired to what was proposed in [9]), but this time with compressed sensing in mind. In traditional compressive temporal imaging systems, the number of sub-frames collapsed in the measurement depends on the number of coded apertures generated by a DMD or discrete movement of a shifting mask. However, for the system proposed in this work, the number of sub-frames depends directly on the sensor parameters $\mathrm{T} = \lceil (t_\alpha + nt_r + t_e)/\Delta t \rceil$. As can be seen in the presented simulation and experimental results, the initial and final frames lack of relevant information since at the beginning and end of the acquisition, the scanlines of each camera are on the edges, so they are disregarded reducing the number of sub-frames to T_{se}. In terms of compression ratio, the number of sampled pixels is determined by $N_{ps} = mnC$ and the reconstructed pixels are $N_{pr} = mn\mathrm{T}_{se}$ for the T_{se} effective sub-frames, leading to a compression ratio of $CR = \mathrm{T}_{se}/C$. When using the 4-camera array, the compression ratio for all simulated datasets was approximately $9X$, while for the experimental results the compression ratio was $14X$. Knowing that we are reconstructing 58 frames from the experimental RS camera array measurements, then if we were to continuously reconstruct the video frames captured at 67 fps we could achieve a reconstructed frame rate of $67 \times 58 = 3,886$ fps.

Optics EXPRESS

6. Conclusion

We have proposed a compressive temporal imaging system based on a camera array of RS-based CMOS cameras. The proposed compressive imaging system exploits the RS dynamics–often considered as a nuisance–as a coding opportunity for the space-time datacube. Since a single RS camera is not capable of performing a good recovery of the datacube due to the poor sampling diversity, we proposed to use an array of differently oriented RS cameras to improve the temporal resolution. By proper modeling and characterizing the RS of an off-the-shelf camera and the camera array geometry, it is possible to generate a forward matrix used to then reconstruct a video from a set of RS snapshots without the need of additional intervention to the imaging system, as often performed by mechanical coded apertures or a DMD in other compressive temporal imaging devices. Our simulation results indicated that using more cameras is beneficial for better reconstruction quality, although the improvement turns to be incremental after 4 cameras. The experimental results obtained with a 2- and 4-camera array finally demonstrated the ability of the proposed system to reconstruct high-speed events with the proper direction and speed, which are hard to recognize solely from the captured RS snapshots. Clearly, the presented results are not comparable in quality to those obtained by state-of-the-art compressive temporal imaging methods, which was never the aim of the presented work. On the other hand, we believe this work shows a novel way to recover high speed events from RS snapshots by exploiting compressed sensing, while also indicating a feasible path for tweaking the RS at the hardware level as the key for achieving a competitive, low-cost, hardware-only compressive temporal imaging device.

Funding. Fondos de Desarrollo de la Astronomía Nacional (QUIMAL190002); Fondo Nacional de Desarrollo Científico y Tecnológico (1181943); Air Force Office of Scientific Research (FA9550-19-1-0293).

Disclosures. The authors declare no conflicts of interest.

References

1. E. Charbon, "Will cmos imagers ever need ultra-high speed?" in *Proceedings. 7th International Conference on Solid-State and Integrated Circuits Technology, 2004.*, vol. 3 (2004), pp. 1975–1980.
2. M. Versluis, "High-speed imaging in fluids," Exp. Fluids **54**(2), 1458 (2013).
3. S. Hertegård, H. Larsson, and T. Wittenberg, "High-speed imaging: applications and development," Logop. Phoniatrics Vocology **28**(3), 133–139 (2003).
4. W. S. Boyle and G. E. Smith, "Charge coupled semiconductor devices," Bell Syst. Tech. J. **49**(4), 587–593 (1970).
5. E. R. Fossum, "Cmos image sensors: electronic camera-on-a-chip," IEEE Trans. Electron Devices **44**(10), 1689–1698 (1997).
6. B. S. Carlson, "Comparison of modern ccd and cmos image sensor technologies and systems for low resolution imaging," in *SENSORS, 2002 IEEE*, vol. 1 (IEEE, 2002), pp. 171–176.
7. D. J. Duke, T. Knast, B. Thethy, L. Gisler, and D. Edgington-Mitchell, "A low-cost high-speed cmos camera for scientific imaging," Meas. Sci. Technol. **30**(7), 075403 (2019).
8. B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2 (IEEE, 2004), p. II.
9. J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *2010 IEEE International Conference on Computational Photography (ICCP)*, (IEEE, 2010), pp. 1–8.
10. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
11. 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**(9), 10526–10545 (2013).
12. R. Pournaghi and X. Wu, "Coded acquisition of high frame rate video," IEEE Trans. on Image Process. **23**(12), 5670–5682 (2014).
13. J. Zhang, T. Xiong, T. Tran, S. Chin, and R. Etienne-Cummings, "Compact all-cmos spatiotemporal compressive sensing video camera with pixel-wise coded exposure," Opt. Express **24**(8), 9013–9024 (2016).
14. N. Antipa, P. Oare, E. Bostan, R. Ng, and L. Waller, "Video from stills: Lensless imaging with rolling shutter," in *2019 IEEE International Conference on Computational Photography (ICCP)*, (IEEE, 2019), pp. 1–8.
15. S. Mendis, S. E. Kemeny, and E. R. Fossum, "Cmos active pixel image sensor," IEEE Trans. Electron Devices **41**(3), 452–453 (1994).
16. Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications* (Cambridge university press, 2012).
17. B. A. Fowler, A. El Gamal, D. X. Yang, and H. Tian, "Method for estimating quantum efficiency for cmos image sensors," in *Solid State Sensor Arrays: Development and Applications II*, vol. 3301 (International Society for Optics and Photonics, 1998), pp. 178–185.

Optics EXPRESS

18. B. Pain and B. R. Hancock, "Accurate estimation of conversion gain and quantum efficiency in cmos imagers," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV*, vol. 5017 (International Society for Optics and Photonics, 2003), pp. 94–103.
19. H. Tian, B. Fowler, and A. E. Gamal, "Analysis of temporal noise in cmos photodiode active pixel sensor," IEEE J. Solid-State Circuits **36**(1), 92–101 (2001).
20. Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," IEEE Trans. Pattern Anal. Mach. Intell. **41**(12), 2990–3006 (2019).
21. X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted-2, 1 minimization with applications to model-based compressive sensing," SIAM J. on Imaging Sci. **7**(2), 797–823 (2014).
22. S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers* (Now Publishers Inc, 2011).
23. X. Yuan, P. Llull, X. Liao, J. Yang, D. J. Brady, G. Sapiro, and L. Carin, "Low-cost compressive sensing for color video and depth," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE Computer Society, Los Alamitos, CA, USA, 2014), pp. 3318–3325.
24. S. Jalali and X. Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," IEEE Trans. Inf. Theory **65**(12), 8005–8024 (2019).
25. X. Yuan, Y. Sun, and S. Pang, "Compressive video sensing with side information," Appl. Opt. **56**(10), 2697–2704 (2017).
26. X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *2016 IEEE International Conference on Image Processing (ICIP)*, (2016), pp. 2539–2543.
27. A. Chambolle, "An algorithm for total variation minimization and applications," J. Math. Imaging Vision **20**(1/2), 89–97 (2004).