SpinCam: High-Speed Imaging via a Rotating Point-Spread Function

Dorian Chan, Mark Sheinin, and Matthew O'Toole
Carnegie Mellon University, Pittsburgh, PA 15213, USA
dychan@andrew.cmu.edu, marksheinin@gmail.com, mpotoole@cmu.edu

Abstract

High-speed cameras are an indispensable tool used for the slow-motion analysis of scenes. However, the fixed bandwidth of any imaging system quickly becomes a bottleneck, resulting in a fundamental trade-off between the camera's spatial and temporal resolutions. In recent years, compressive high-speed imaging systems have been proposed to circumvent these issues by optically encoding the signal and using a reconstruction procedure to recover a video. Our work proposes a novel approach for compressive high-speed imaging based on temporally coding the camera's point-spread function (PSF). By mechanically spinning a diffraction grating in front of a camera, the sensor integrates an image blurred by a PSF that continuously rotates over time. We also propose a deconvolution-based reconstruction algorithm to reconstruct videos from these measurements. Our method achieves superior light efficiency and handles a wider scene class than prior methods. Also, our mechanical design yields flexible temporal resolution that can be easily increased, potentially allowing capture at 192 kHz—far higher than prior works. We demonstrate a prototype for various applications, including motion capture and particle image velocimetry (PIV).

1. Introduction

The world is full of high-speed events that are imperceptible to both the human eye and most conventional cameras (*e.g.*, the flicker of electric lights or the motion of fast-moving objects). The traditional approach to capturing such events is using slow-motion photography, *i.e.*, use a camera to shoot images at a high frame rate and display the results at a lower frame rate. Slow-motion photography is often used for artistic effects in film and as a measurement tool in scientific applications. However, despite significant advancements in sensor technology, conventional high-speed cameras tend to be expensive, bulky, and have high throughput requirements for recording visual content at both high frame rates and high spatial resolutions. These hardware limitations set an upper limit to the cameras' capture rates.
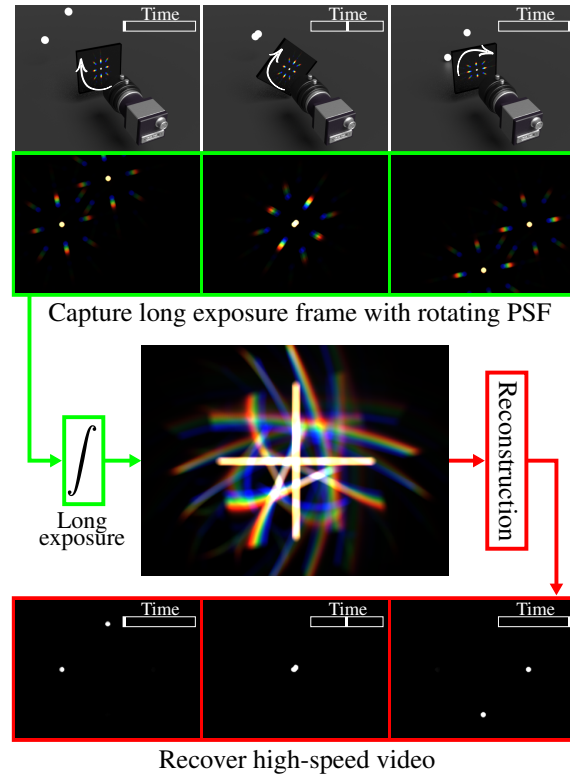

Figure 1. **Principle of high-speed imaging with a rotating diffraction grating.** A camera captures a long exposure image of a moving scene. During the camera's exposure period, the diffraction grating placed in front of the camera continuously rotates to generate a time-varying point-spread function (PSF) that encodes scene dynamics into a single frame. Solving a deconvolution problem yields a high-speed video from the captured frame.

Compressed sensing provides a framework to circumvent such issues. Since videos are highly compressible, it is possible to recover a video from a single snapshot by (i) compressively recording spatiotemporal visual information onto a sensor and (ii) solving a reconstruction problem to recover the video using priors on the visual content. Among the large body of work in compressive high-speed imaging, most techniques temporally modulate the light incident on sensor pixels [48, 12, 20, 22, 9, 34, 32, 35, 14, 16, 19,

13, 49], either through the use of a spatial light modulator (SLM) or novel sensors capable of turning pixels on or off. Alternatively, by using spatial-multiplexing optics (*e.g.*, diffusers, diffraction gratings) to distribute light from any scene point to many camera pixels, it is possible to perform high-speed imaging by using a rolling shutter [2, 50, 26], a reduced region-of-interest (ROI) [39], or a line sensor [40].

Our work proposes a fundamentally different approach to compressive high-speed imaging. Similar to past works that make use of spatial-multiplexing optics, we propose to distribute light across a sensor by using a dual-axis diffraction grating, which diffracts light horizontally and vertically (see Fig. 1). Unlike prior work, however, we opt to change the point-spread function (PSF) *during* the camera's exposure period. Specifically, we use a motor to mechanically rotate the PSF at high speeds, record a single global-shutter frame, and solve a simple deconvolution-based reconstruction procedure to reconstruct scene dynamics. Please refer to Fig. 1 for an overview of the process.

Our approach, like past work using spatial-multiplexing optics [2, 50, 40, 39], is primarily targeted towards spatiotemporally sparse scenes. This imaging regime includes motion capture and particle-image velocimetry (PIV), which see broad application in domains like biomechanics [29], sports science [31], animation and filmmaking [38], virtual and augmented reality [5], aerodynamics [47], fluid mechanics [36], biomedical engineering [18], oceanography [7], and combustion and engine research [28]. We demonstrate our approach for motion capture and PIV in Sec. 6. Other potential applications for spatiotemporally sparse video include high-speed 3D scanning [24] and capturing the fast flicker of electric bulbs [41].

Our approach combines four key advantages over existing methods targeting spatiotemporally sparse scenes:

(I) Superior light efficiency. In many prior methods, the scene light is dispersed using specialized optics to reach many camera pixels, but only a *small subset* of camera pixels collect light at each timestamp [2, 40, 39, 50]. Thus, most light is lost by falling outside the 'active' pixels. Conversely, in our work, the signal from different timestamps is encoded by different temporal PSFs, and the entire sensor is left exposed, yielding far superior light efficiency. We show this advantage in simulated experiments in Sec. 3.

(II) Easily increasable temporal resolution. In most prior works, temporal resolution (*i.e.*, FPS) is fundamentally limited by the speed of the respective technology (*e.g.*, the modulation speed of a SLM/sensor [12, 22, 34, 32, 14, 49], or the read-out speed of a camera [2, 40, 39, 50, 9]). In contrast, the temporal resolution of our approach is determined by the mechanical rotation speed and thus can be easily increased by rotating the PSF faster. Thus, we show that our approach can resolve a LED flickering at 96 kHz.

(III) A wider class of supported scenes. Prior works using spatial-multiplexing optics allocate a fixed amount of pixels for each timestamp (*e.g.*, a single rolling shutter row) [2, 40, 39, 50]. This limits the amount of information recoverable per frame, imposing a spatial sparsity constraint on each high-speed frame individually. Instead, our solution multiplexes the signals from different timestamps together, yielding a sparsity constraint over the *entire video*. Our method can thus handle temporally sparse but spatially dense scenes, verified in simulated and real experiments.

(IV) Simple hardware. While many prior works require specialized optics (*e.g.*, SLMs [12, 34, 49], piezoelectric stages [20, 16, 19]) or specialized sensors (*e.g.*, coded exposure pixels [22, 14, 32]), our setup can work with any global-shutter camera combined with simple low-cost hardware that is readily available at your local hobby store. We include a DIY guide as part of our supplementary materials.

2. Related work

2.1. Spinning Disks & Revolving Mirrors

Mechanical spinning disks and revolving mirrors have been used to measure transient phenomena for centuries. In 1834, Wheatstone [51] estimated the velocity of electricity traveling through long copper wires by observing the sparks generated at the ends. Using a revolving mirror, Wheatstone observed a 1 μs time difference, and concluded the velocity of electricity to be 288,000 miles per second. Fizeau and Foucault [25] later measured the speed of light through various apparatuses based on revolving mirrors and spinning cogwheels. For example, in 1848-49, Fizeau measured the transit time of light from a source, to a mirror, and back to the source by using a rotating wheel to modulate the light.

Though rare in modern consumer electronics, modulating light mechanically with a spinning element is both faster and lower cost than most modern digital modulation solutions. For example, digital micromirror devices (DMDs) only operate up to 30 kHz, whereas spinning disks can offer rates up to 2.4 MHz [10]. As a result, coded disks have been used for single pixel imaging at rates up to 72 FPS [43, 42, 10], and revolving coded mirrors have been used to reconstruct 1400 frames from a single shot [23]. A spinning camera was used to recover an 11-frame video through scattering media by disambiguating the resulting rotated speckle patterns [44], while a rotating camera coupled with a polarizer generalized the optical-flow equation to include polarization cues [46]. An event camera can be combined with a spinning polarizer for fast, accurate shape-from-polarization [27]. This work explores the use of spinning elements for compressive high-speed video.

2.2. Pixel-wise Coded Exposure

Coded exposure photography involves temporally modulating the light incident on a sensor, which has long been

used to capture dynamic events [4]. Raskar *et al*. [33] used a liquid-crystal panel to rapidly open and close the shutter of a camera; the motion blur associated with fast-moving objects could then be reliably removed through deconvolution. This methodology was later extended for recovering full high-speed videos of periodic [48] and generic scenes [13].

A popular approach to compressive high-speed imaging involves pixel-wise exposure control [12, 9, 34, 32, 35, 22, 14, 20, 16, 19]. In practice, the codes are often binary, and the corresponding reconstruction techniques rely on various priors to reconstruct videos. Example priors include assuming the scene produces periodic motion [48] or satisfies the brightness constancy assumption [34], and might rely on dictionary learning [12, 13] and neural networks [22, 14, 30] to aid the reconstruction procedure.

These techniques lack the **(II)** temporal resolution and **(IV)** hardware complexity advantages of our approach since they typically require specialized hardware (*e.g.*, SLMs [12, 34, 49], piezoelectric stages [20, 16, 19]) or specialized sensors [22, 14, 32]) which can be expensive, difficult to acquire, and still be limited in frame rate. Moreover, as shown in the supplementary materials, since these methods do not spread light across multiple pixels, these systems struggle with spatially sparse scenes, which appear in many practical applications of high-speed imaging [40, 39].

2.3. Spatial-Multiplexing Optics

An alternative approach to capture space-time volumes (videos) involves spatial multiplexing. By placing an optical element in front of a sensor or camera (*e.g.*, a diffuser or diffraction grating), the light from a single point in the scene spreads to multiple camera pixels according to the optical system's PSF. When combined with a rolling shutter camera that samples individual sensor rows at fast rates [2, 50, 26], it is possible to reconstruct high-speed videos. Alternatively, a limited region-of-interest (ROI) or 1D sensor can also be used to sample measurements at fast rates [40, 39]. By integrating diffraction patterns generated by time-varying coded masks, it is even possible to reconstruct 4D volumes of spacetime from a single snapshot [49].

Similar ideas have been used in combination with streak cameras, an optoelectronic device used to measure ultra-fast light phenomena at femtosecond time scales. A streak camera spatially multiplexes the scene's temporal response, by deflecting photoelectrons to different points on a sensor as a function of time. While designed to be a line sensor, it is possible to encode 2D spatial information by incorporating a tilted microlens [45] or pinhole array [11]. In Gao *et al*. [8], the 2D scene at different timestamps is coded with a fixed mask, sheared a temporally varying amount, and then overlayed onto the 2D signal captured by a streak camera.

These methods lack the **(I)** light efficiency and **(III)** density advantages of our approach—they typically rely

(a) Captured image
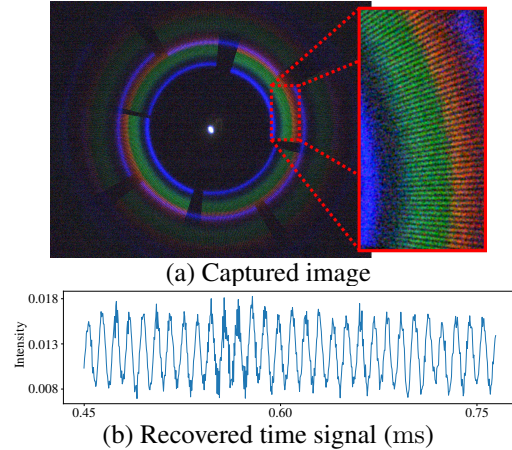

(b) Recovered time signal (ms)

Figure 2. **High-speed reconstruction of a flickering LED.** The LED in this experiment flashes at 96 kHz, as driven by an Arduino. **(a)** The diffraction image encodes the high-frequency patterns of the LED (zoom into inset). **(b)** This temporally-varying signal can be directly extracted from the image to recover the intensity pattern of the LED. A frequency analysis of the recovered time signal indicates the system's ability to resolve frequencies up to 96 kHz. Given that the Nyquist frequency is half that of the sample rate, the effective framerate for the system is therefore over 192,000 FPS.

on a small subset of camera pixels to reconstruct each frame, "wasting" light and limiting recoverable scene density. Furthermore, their capture speed is limited by the maximum speed of the hardware used (*i.e.*, line-scan mode [40], ROIs [39] or rolling shutter [2, 50]). Thus, these methods lack the ability to **(II)** easily increase temporal resolution.

3. High-Speed Imaging using a Time-Varying Point-Spread Function

Consider a scene that consists of a single bright LED in a dark room, where the LED is configured to pulse rapidly over time. Unless it is faster than the blinking LED, a normal camera forms an image of the LED, but provides no information about the LED's temporal dynamics.

Our proposed approach to imaging high-speed events involves two steps. First, following prior work on diffraction imaging [40, 39], we place a diffraction grating in front of the camera, which blurs the scene with a rainbow PSF. The light from the LED is therefore spread over many additional camera pixels. Second, we propose to rotate the diffraction grating such that the corresponding PSF changes over time. This produces measurements that spatially encode temporally varying signals, which can be used to analyze extremely high-speed events (see Fig. 2 for an example).

In this section, we describe the image formation model and reconstruction process for recovering a high-speed video from an image captured with a generic time-varying PSF. We then analyze the special case of using a rotating PSF for capturing images.

(a) Recovered brightness of time-varying light source
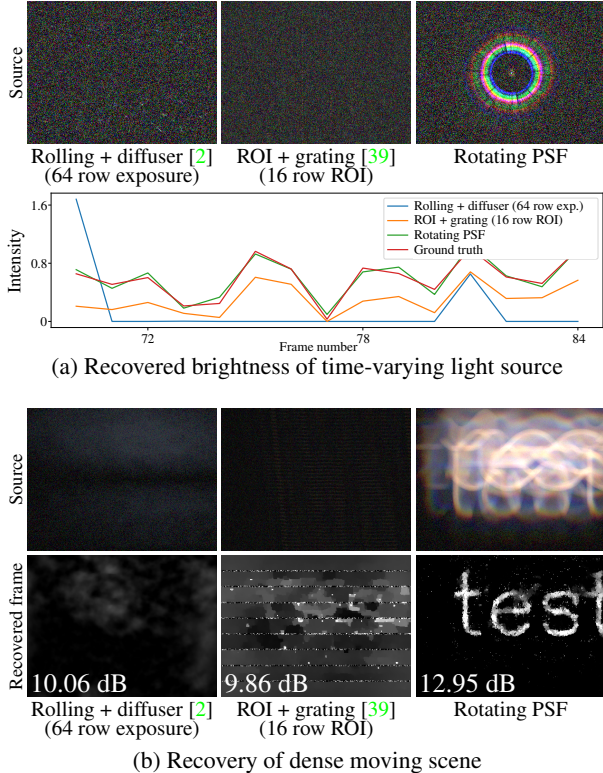

(b) Recovery of dense moving scene

Figure 3. **Light efficiency experiment.** In the presence of Poisson and read noise, we simulated the measurements captured by a Canon 5D Mark III sensor [6] with ISO 800. **(a)** We simulate a dim point light source with a time-varying intensity profile. **(b)** We simulate the word *"test"* moving across the scene. In both cases, our approach is able to much more accurately resolve the signal, compared to a rolling shutter and diffuser as well as a ROI and grating. We theorize that this is due to the fact that our approach captures more light for every time step, and that the requisite de-multiplexing minimizes the effects of noise.

3.1. Image Formation Model

As illustrated in Fig. 1, for a camera where the PSF changes with time, the forward model can be written as:

$$\mathbf{I}(x, y) = \int_{t=t_s}^{t=t_s+E} (\mathbf{k}(t) * \mathbf{s}(t)) (x, y) \, dt, \qquad (1)$$

where $\mathbf{I}(x, y)$ is the captured image, E denotes the exposure time of the camera, and t_s denotes the time when the camera started exposing. The symbol $*$ denotes the spatial convolution operator between the scene $\mathbf{s}(t)$ and the PSF $\mathbf{k}(t)$ at time t. When discretized to N_E time steps from t_s to $t_s + E$, this can be rewritten in matrix form as:

$$\mathbf{I} = \mathbf{M}\mathbf{s} \qquad (2)$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_{N_E} \end{bmatrix} \qquad (3)$$

where $\mathbf{I} \in \mathbb{R}_+^{N_x N_y}$ is the vectorized representation of the captured image, $\mathbf{s} \in \mathbb{R}_+^{N_x N_y N_E}$ is the vectorized representation of the high-speed scene, and $\mathbf{C}_k \in \mathbb{R}_+^{N_x N_y \times N_x N_y}$ represents the convolution with the kernel at timestep k.

3.2. Reconstruction Process

Natural scenes contain sparse gradients, and many target applications for high-speed imaging require the tracking of just sparse points. As a result, we recover a high-speed video with an ℓ_1 sparsity prior and an anisotropic total variation prior, solving the following optimization problem:

$$\operatorname*{argmin}_{\mathbf{x}} \quad \|\mathbf{I} - \mathbf{M}\mathbf{x}\|^2 + \mathcal{I}_{\mathbb{R}_+}(\mathbf{x}) + \lambda_{sparsity}\|\mathbf{x}\|_1 +$$
$$\lambda_{dx}\|\mathbf{D}_x\mathbf{x}\|_1 + \lambda_{dy}\|\mathbf{D}_y\mathbf{x}\|_1 + \lambda_{dt}\|\mathbf{D}_t\mathbf{x}\|_1 \qquad (4)$$

where $\mathcal{I}_{\mathbb{R}_+}$ denotes a non-negativity prior, $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_t$ compute forward finite differences along the x, y, and t dimensions respectively of the high-speed video $\mathbf{x}$, and $\lambda_{sparsity}$, λ_{dx}, λ_{dy}, and λ_{dt} weigh the regularization terms.
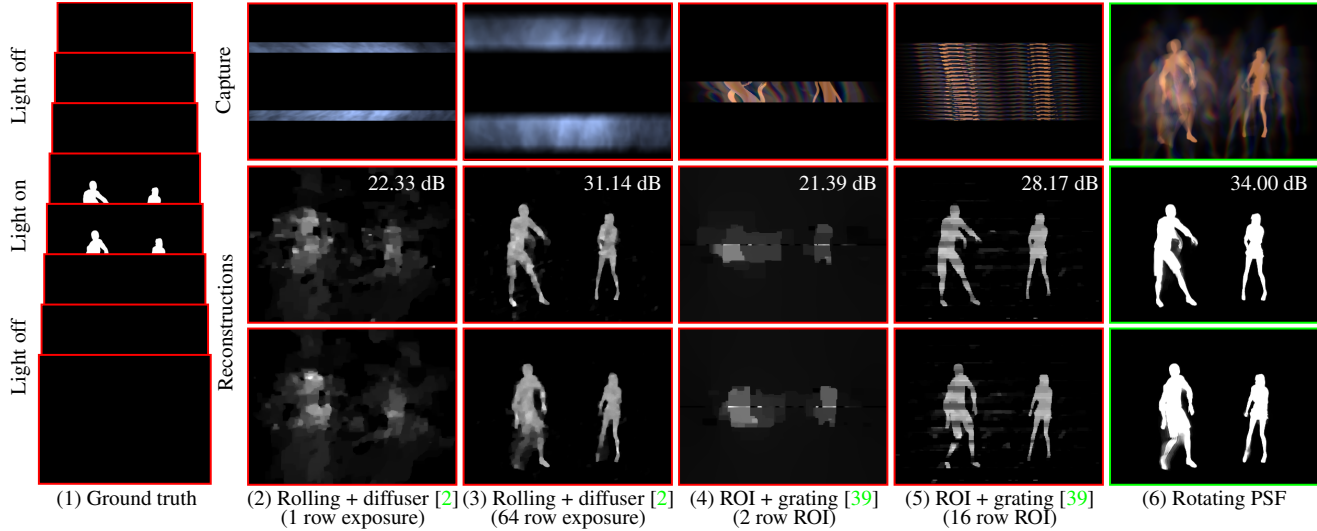
3.3. Analysis

Varying the PSF over time results in fundamentally different properties compared to prior approaches, which we explore in the rest of this section.

Higher light efficiency. Our method, like other prior compressive video works [2, 50, 40, 39], relies on PSFs with broad spatial supports. A broad spatial support spreads the signal from a single point across multiple pixels. Recent approaches record just a subset of these pixels at each timestep [2, 50, 40, 39]. Because fewer pixels encode each frame and light is lost, assuming a standard noise model with Poisson shot noise and additive Gaussian noise, the reconstruction's signal-to-noise ratio is proportionately decreased. In contrast, under a time-varying PSF, the entire frame is always recorded, so a much larger fraction of the PSF (and therefore light) is captured for every scene point during capture. In addition, the signals from different timestamps are mixed together in the final capture, which could improve the quality of the reconstructions under read noise [37]. Thus, this overall increase in light efficiency leads to better performance in resolving noisy scenes. We show this experimentally in Fig. 3 on two simulated cases: a point light source with time-varying intensity, and a denser moving scene. We find that our approach robustly recovers these two scenes, even under low light conditions.

A wider class of scenes. Many past approaches achieve high-speeds by allocating a number of pixels per frames that is much lower than the total number of frame pixels [2, 50, 40, 39]. This causes each frame to undergo lossy compression, yielding a sparsity constraint per frame. For example, for a rolling shutter with exposure set to the row rate, each frame is encoded onto just a single sensor row.

(a) Simulated scene:

| (1) Ground truth | (2) Rolling + diffuser [2] (1 row exposure) | (3) Rolling + diffuser [2] (64 row exposure) | (4) ROI + grating [39] (2 row ROI) | (5) ROI + grating [39] (16 row ROI) | (6) Rotating PSF |

(b) Real scene:

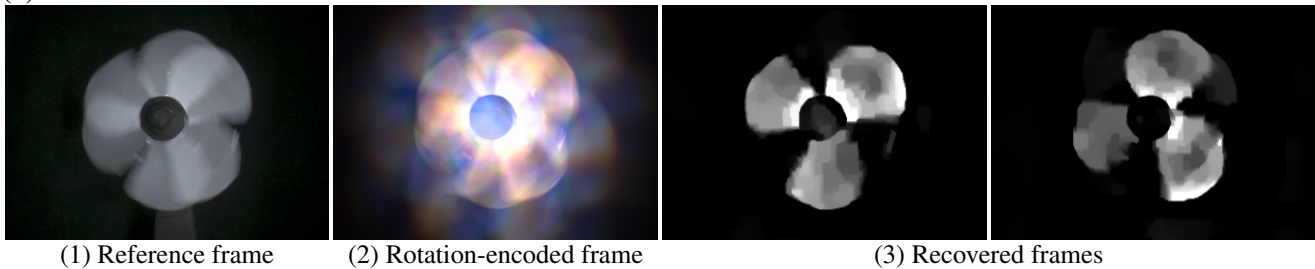| (1) Reference frame | (2) Rotation-encoded frame | (3) Recovered frames |

Figure 4. **Reconstructing temporally sparse, but spatially dense scenes. (a)** A simulated scene of two people dancing. The video is dark for the entire sequence, except for a short sequence of fast-moving frames where the dancers are strobed. **(a2)-(a3)** We implement a setup based on a dual rolling shutter combined with a diffuser [2]. In (a2), with the exposure time set to match the row rate of the shutter, the reconstruction is unable to well-resolve the dense scene. In (a3), with the exposure time set to 64 times the previous value, the reconstruction is able to recover the dancers, at the cost of temporal resolution [2]. **(a4)-(a5)** We implement an approach based on fast temporal sampling by using a ROI and a diffraction grating [39]. In (a4), the ROI consists of just two rows, resulting in low quality results. In (a5), by using more bandwidth and exposing 16 rows, such an approach can get better results, but this reduces the recoverable framerate by 8 times in practice. **(a6)** In contrast, our methodology recovers the dense scene much more accurately compared to (a2) and (a4)—the signal is encoded over a much larger portion of sensor. This does not require increasing the exposure time and therefore decreasing the temporal resolution as in (a3), or increasing the requisite bandwidth (a5). **(b)** A real world scene, where a spinning fan is strobed twice during the camera exposure at unknown times. Our approach is able to disambiguate the fan rotation at each strobe time.

In contrast, our method allocates the entire frame to encode a plurality of video frames belonging to a fixed time interval. Therefore, our method assumes sparsity in a more general sense, namely sparsity over the entire fixed time interval. The encoded fixed time interval can be either spatially sparse and temporally dense, which fits most prior-work assumptions, or *temporally sparse but spatially dense*, which prior approaches do not consider due to the fixed bit-rate per frame imposed by their low-dimensional encoding.

To illustrate how this is an advantage, consider the case of a dense scene that is entirely dark, except for a single time instant with a flash of light. If sampling the scene using a rolling-shutter or line sensor, the flashed frame will be encoded onto just a single row. In contrast, a time-varying

PSF encodes the flashed frame on the entire frame, wasting no resources encoding the dark frames. Therefore, in this case, a time-varying PSF enables a better reconstruction.

As shown in Fig. 4, handling temporally sparse, but spatially dense scenes makes our approach very useful under strobed lighting. As shown in real and simulated experiments, a rotating PSF is able to resolve the content within each individual strobe at high spatiotemporal resolution, without any constraints on scene structure between strobes.

4. Using a Rotating PSF

Sec. 3 described the encoding and reconstruction of high-speed video using a time-varying PSF. While a time-varying PSF can be implemented in a number of ways,
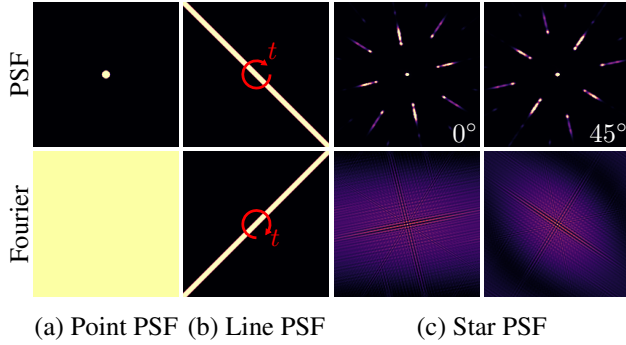
(a) Point PSF (b) Line PSF   (c) Star PSF

Figure 5. **The PSF and spatial frequency coding of time**. A rotating PSF effectively codes the signal from different timesteps into a rotating set of spatial frequencies in the captured image. For instance, a rotating line PSF encodes different timestamps into different lines in Fourier space. The star PSF from a double-axis diffraction grating behaves as a mixture of the point and line cases. Its Fourier transform contains bright line components that provide temporal signal, but remains broadband for spatial resolution.

for the sake of cost and speed, we simply rotate a PSF-generating optical element over the course of the camera exposure. This results in a rotation of the corresponding PSF. In this particular case, $\mathbf{k}(t)$ can now be rewritten as $\mathbf{k}(t) = \mathbf{R}(\mathbf{k}, \theta_t)$, where $\mathbf{k}$ is the stationary PSF of the rotating optical element, θ_t is the angle of the element at time t, and $\mathbf{R}(\mathbf{k}, \theta_t)$ denotes a rotation of $\mathbf{k}$ by θ_t. In the rest of this section, we analyze the properties of such a rotating setup.

PSF performance. The selected PSF affects the spatio-temporal reconstruction quality of every video frame. Because each time step is convolved with a rotated PSF, the PSF's rotated Fourier transform determines which parts of the spectrum will encode each time step.[1] To illustrate, an ideal thin lens having a perfect impulse PSF will have perfect spatial resolution because its Fourier transform contributes values in all spectrum frequencies (Fig. 5(a)). It will, however, have poor temporal resolution because every other timestamp will *also* contribute values across the entire spectrum, making them impossible to distinguish.

In contrast, consider a line PSF (Fig. 5(b)). In the Fourier domain, the PSF remains a line that goes through the origin. When the line PSF rotates, its Fourier spectrum rotates correspondingly, such that each time step is encoded into a different line in Fourier space, without interfering with values from previous timestamps. As a result, the signal at each time step can be easily reconstructed simply by extracting its corresponding line from Fourier space. However, because only a single direction of spatial frequencies is available for reconstruction for every time step, the re-

---

[1]The Fourier transform of the convolution of two functions is the element-wise product of their Fourier transforms. If a function is rotated by θ, its Fourier transform is the Fourier transform of the original function, rotated by θ.

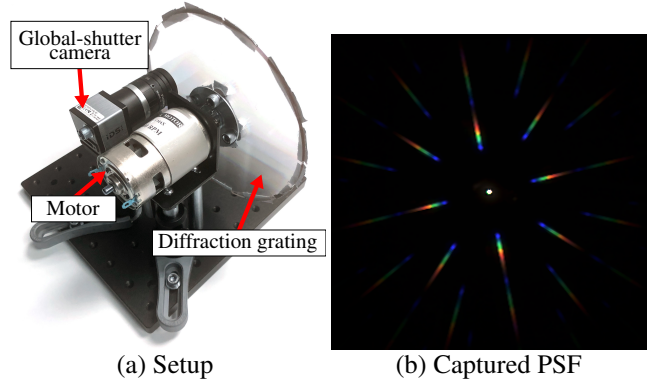(a) Setup   (b) Captured PSF

Figure 6. Image of our hardware setup, consisting of a global-shutter camera and a circular diffraction grating attached to a 20,000 RPM motor.

sulting recovered frame will have poor spatial resolution.

Inspired by these two cases, we use a double-axis diffraction grating for the experiments shown in this work, whose corresponding PSF is shown in Fig. 6(b). This star PSF contains a bright DC spot at the center, intuitively providing spatial resolution. Simultaneously, the PSF also contains line-shaped rainbow lobes that provide temporal information. We visualize this phenomena in Fig. 5(c).

Temporal resolution. The *temporal* resolution of our system is fundamentally determined by the amount of temporal variation of the PSF between different timestamps, or namely, the minimum angle required to reliably distinguish between two different rotations of the PSF. As a result, the temporal resolution can be increased either by increasing the pixel resolution, the spatial spread of the PSF, or the rotation speed. But since our grating PSF is roughly $90°$ periodic, rotating more than $90°$ during the exposure causes temporal ambiguity and degrades the reconstruction quality.

In the supplement, we further analyze PSF performance from the mutual coherence perspective. We also discuss how a time-varying PSF behaves as a form of coded exposure in frequency, as shown by the above Fourier analysis.

5. Implementation

Hardware. An image of our hardware setup is given in Fig. 6. To generate a usable PSF, we cut a 531-lines per millimeter double-axis diffraction grating sheet into a circular disk of radius 6 inches. To ensure that the sheet stayed flat, we taped it onto a laser-cut acrylic disk with a radius of 6 inches. We attached this element to a Puly DC 775 Motor with a shaft coupling to rotate it. We place this entire rotating setup $5\,\mathrm{mm}$ in front of the camera's lens. We used a UI-3070CP-C-HQ R2 camera, fitted with a $9\,\mathrm{mm}$ $f/1.4$ lens (Fujinon HF9HA). We also placed a second identical camera beside our system to capture intermittent reference

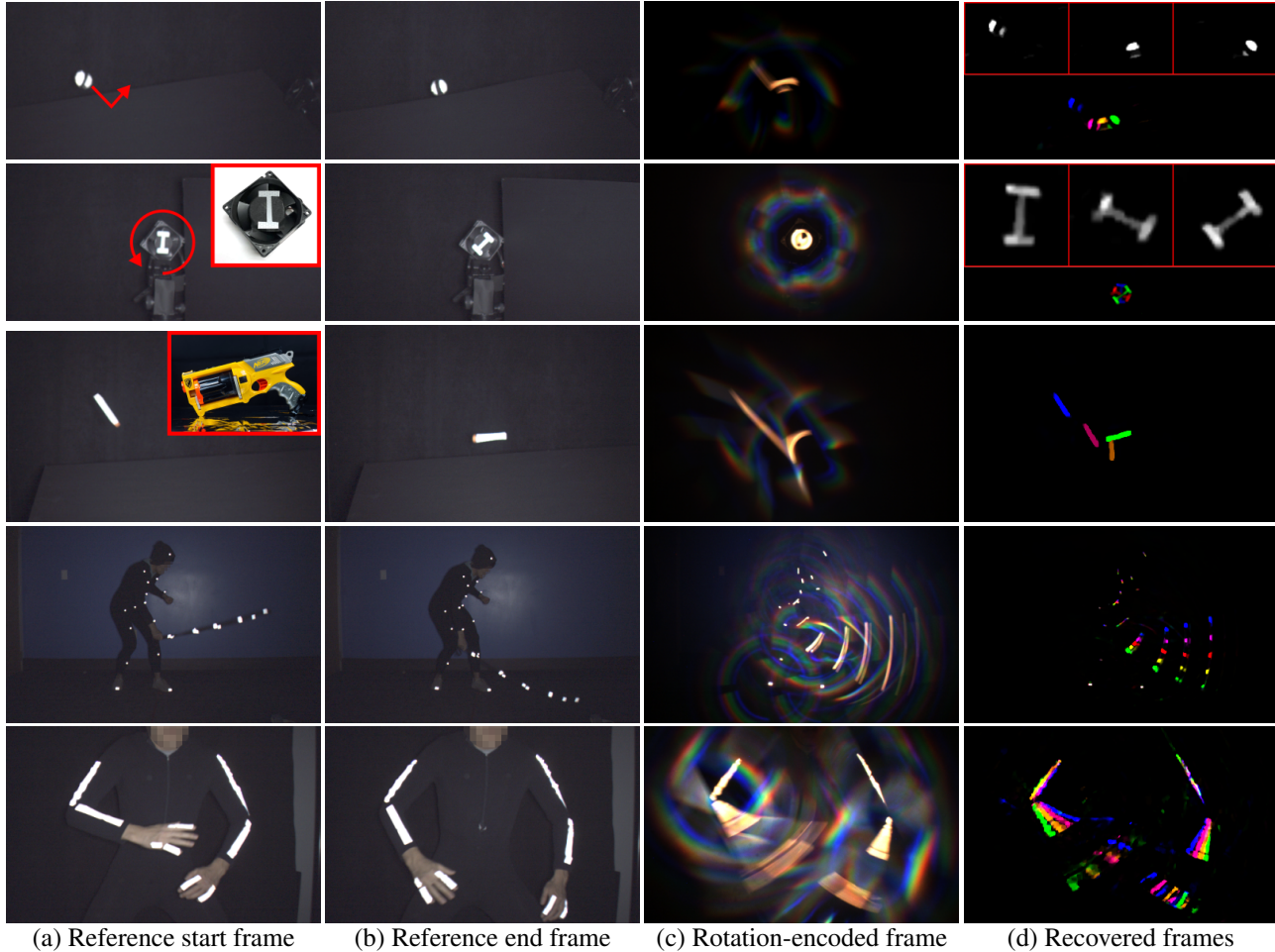|			
(a) Reference start frame	(b) Reference end frame	(c) Rotation-encoded frame	(d) Recovered frames

Figure 7. High-speed reconstruction of fast moving objects. **Row 1:** A ball with a black stripe bounces and spins. **Row 2:** An I rotates 120°. **Row 3:** A toy bullet bouncing off a board. **Row 4, 5:** Our setup is capable of tracking the movements of both sparse points and dense marker lines for high-speed motion capture.

frames for visualization. Please refer to the supplement for a DIY guide.

Software. To solve Eq. (4), we used a PyTorch implementation of FISTA [3], combined with an approximate proximal operator [15] to handle the anisotropic total variation prior. We ran 10000 iterations of FISTA for each result shown. For most results, we set $\lambda_{sparsity} = 0$, $\lambda_{dx} = \lambda_{dy} = 1 \times 10^{-3}$, and $\lambda_{dt} = 1 \times 10^{-3}$, but we vary the strength of these terms depending on the brightness of the background and the speed of the motion. We list the parameters for each result and provide sample code in the supplemental.

Calibration and capture. To calibrate the time-varying PSF of our imaging system, we process a stationary image of a small retroreflector or LED, and then computationally rotate this image to create the dictionary required for Eq. (4)—please see the supplemental for more details.

For every capture, we need to identify the range of angles that are rotated over during the camera exposure in order to solve Eq. (4). This could be done by placing an LED at a known location, and reconstructing which PSFs are present around it. Alternatively, a controllable motor or trigger system could be used to directly provide the requisite angles. In our work, we opt for a unstructured approach, that simply relies on signal being present at the start and end of the camera exposure. Before capture, we first calibrate the motor to rotate roughly 90° over the course of the exposure. Then, we simply perform a coarse reconstruction of Eq. (4) using the captured image and a dictionary that spans all 360°, and select the best matching 90° for a finer reconstruction. Please see the supplemental for an in-depth discussion.

6. Results

For our real-world reconstructions, we downsampled the input images to 300×400 pixels and reconstructed $N_E = 146$ frames for every input. For the simulated results (Fig. 3, Fig. 4), we instead used $N_E = 150$ to better match

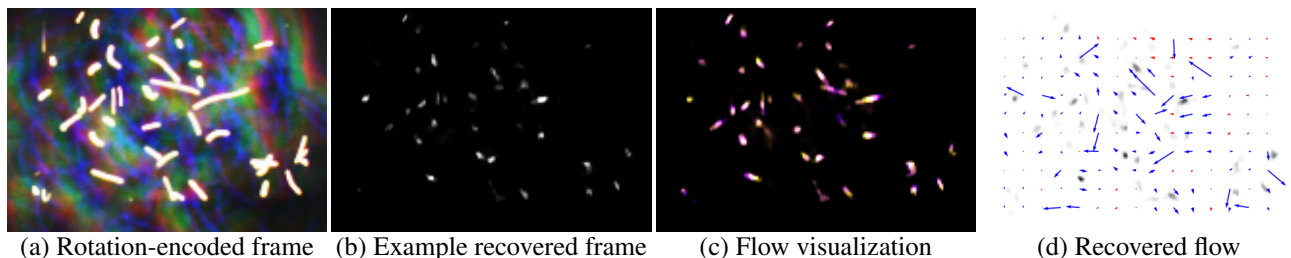(a) Rotation-encoded frame (b) Example recovered frame (c) Flow visualization (d) Recovered flow

Figure 8. **High-speed particle image velocimetry.** Following Sheinin *et al.* [39], a magnetic stirrer is placed in a tank seeded with tracer particles. **(a)** Captured diffraction image. **(b)** Recovered frame using our reconstruction procedure. **(c)** We visualize a $17.33\,\text{ms}$ interval by overlaying the recovered frames. **(d)** Using OpenPIV [17], we recover the flow between the start and end frame of (c).

a dual rolling shutter setup. In this document, we use color to code a video into a single image; we refer the reader to our supplemental materials for directly viewing the recovered videos. The camera exposure was $22.178\,\text{ms}$, yielding an effective framerate of 6583.10 FPS.

We demonstrate our methodology on fast scenes on multiple examples (Fig. 7)—a bouncing ball, an "I" spinning on a fan, toy bullets bouncing off a board, and letters spinning and falling. In the last row of Fig. 7, we applied our technique to marker-based motion capture. Our technique readily captures a wide variety of marker shapes. In Fig. 4(b), we show a temporally sparse, but spatially dense scene of a fan under strobed illumination.

As shown by Sheinin *et al.* [39], low-cost, high-speed imaging could help measure the movement of fluids in particle image velocimetry [1, 21, 52]. In short, small tracer particles are placed into the medium, which can then be tracked to visualize the flow. We apply our setup to this task in Fig. 8, imaging a tank filled with water and a magnetic stirrer. To maximize the visible movement, we increased the exposure time of our camera to $100\,\text{ms}$ and proportionately decreased the speed of the motor.

Finally, we push the temporal resolution of our approach by imaging an LED pulsed at $96\,\text{kHz}$. For this particular result, we used a $16\,\text{mm}$ focal length lens (Edmund Optics #86-571) to increase the scale of the PSF, and rotated the motor at roughly $15,000$ RPM. At resolution 2048×1542 pixels, our methodology is capable of easily resolving this signal as shown in Fig. 2, demonstrating $192\,\text{kHz}$ capture.

7. Limitations

Temporal unmixing. Rolling shutter or ROI-based approaches [2, 50, 40, 39] code the signal from each timestamp onto a unique set of measurements that only record the signal at that time. In contrast, with a rotating PSF, the signal from different timestamps is mixed together and requires unmixing using Eq. (4). This unmixing can yield degraded results when there are significant nonidealities, *e.g.*, a poorly calibrated PSF, major spatial variation of the PSF, sizable background signal, or when the scene is too dense.
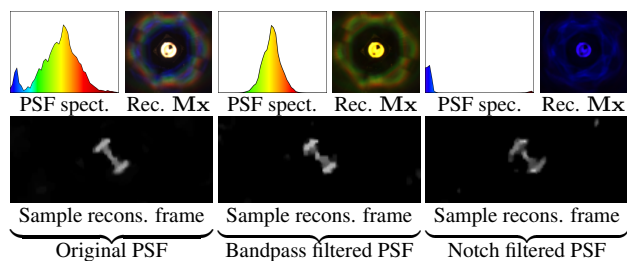

Figure 9. Reconstruction sensitivity to PSF spectrum. We reconstruct Fig. 7, Row 2 with digitally augmented versions of the original real measured PSF, to simulate a PSF calibrated under incorrect lighting or from a different material. The reconstruction quality is fairly robust to these PSF variations.

In these cases, the signal from one timestamp may bleed into the reconstructions for other timestamps, causing a loss of temporal resolution.

Saturation. Our method can be especially prone to sensor saturation for several reasons. First, the use of a longer camera exposure may saturate those camera pixels which received scene light over a significant portion of the exposure duration. Secondly, a bright flash of light at a single timestamp could saturate the entire sensor, corrupting the whole video sequence, while only a single row (and therefore timestamp) would be saturated for a rolling shutter or ROI. Thirdly, our grating's PSF contains a 'DC spot' that is much brighter than the rainbow streaks. Therefore, well-exposed rainbow streaks yield a large DC component that might saturate camera pixels. This latter problem could be mitigated by using a higher-quality grating.

Spectrum. The PSF of an object, when viewed through a diffraction grating, depends on its reflectance spectrum, which itself depends on the light source spectrum and object material. Our approach assumes that scenes consist of a single reflectance spectrum so that this PSF is known and fixed for every point in the scene. This assumption makes our approach well-suited for vision applications like motion capture or PIV, which use specialized light sources and markers. In these cases, only slight deviations in PSF occur, which our method can handle (see Fig. 9). However, a

more general scene may consist of many diverse materials and light sources. This results in many different reflectance spectra and corresponding PSFs at the sensor, restricting the applicability of our method.

To address this limitation, the single reflectance spectrum assumption can potentially be relaxed, albeit with tradeoffs. For instance, the reconstruction problem in Eq. (4) can be expanded to explicitly account for these multiple PSFs following Sheinin *et al*. [39], at the cost of extra computational complexity that scales with the number of possible spectra. Alternatively, a less wavelength-dependent optic (*e.g.*, a simple diffuser [2]) could be used, such that the PSF does not significantly vary for different wavelengths. However, this loses the additional degrees of information encoded in the different PSF colors.

8. Conclusion and Future Work

We demonstrated a novel compressive high-speed video methodology that relies on temporally coding the camera's PSF using a spinning diffraction grating. Our approach can turn ordinary cameras, regardless of their designed speed, into high-speed sensors for spatiotemporally sparse scenes. Compared to prior work, our method can tackle different scene types, is more light efficient, and its temporal resolution is easily increased.

Our prototype relies on a PSF created by a low-cost diffraction grating. Better PSFs, including ones that are optimized using an end-to-end framework [14, 22, 30], would likely produce superior results. Moreover, there are likely more sophisticated mechanical ways to temporally modulate the PSF, potentially providing better performance. For instance, a fast stepper motor that rotates the PSF at discrete steps, synchronized with the reconstruction resolution, could help minimize the mismatch between the discretized model in Eq. (4) and actual measurements. Additionally, other forms of temporal modulation beyond rotation could also prove useful, such as axially translating the spatial-multiplexing optic to scale the PSF, or using a spatial-light modulator to create arbitrary programmable PSFs.

Additionally, better reconstruction procedures that leverage optical flow [34], dictionary learning [12, 13], or deep learning [14, 22, 30] could significantly improve quality over our naive implementations. Finally, our method is orthogonal to previous compressive video approaches mentioned in Sec. 2; through novel combinations of these techniques, we believe it may be possible to spin out reconstructions at even higher spatial and temporal resolutions.

References

[1] Ronald J Adrian and Jerry Westerweel. *Particle image velocimetry*. Number 30. Cambridge university press, 2011. 8

[2] Nick Antipa, Patrick Oare, Emrah Bostan, Ren Ng, and Laura Waller. Video from stills: Lensless imaging with rolling shutter. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2019. 2, 3, 4, 5, 8, 9

[3] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 7

[4] Joyce Bedi, Douglas Collins, and Harold Eugene Edgerton. *Seeing the unseen: Dr. Harold E. Edgerton and the wonders of Strobe Alley*. mIt press, 1994. 3

[5] Jacky CP Chan, Howard Leung, Jeff KT Tang, and Taku Komura. A virtual reality dance training system using motion capture technology. *IEEE transactions on learning technologies*, 4(2):187–195, 2010. 2

[6] Roger N. Clark. Clarkvision.com. 4

[7] P Doron, L Bertuccioli, J Katz, and TR Osborn. Turbulence characteristics and dissipation estimates in the coastal ocean bottom boundary layer from piv data. *Journal of Physical Oceanography*, 31(8):2108–2134, 2001. 2

[8] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014. 3

[9] Jinwei Gu, Yasunobu Hitomi, Tomoo Mitsunaga, and Shree Nayar. Coded rolling shutter photography: Flexible space-time sampling. In *2010 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2010. 2, 3

[10] Evgeny Hahamovich, Sagi Monin, Yoav Hazan, and Amir Rosenthal. Single pixel imaging at megahertz switching rates via cyclic hadamard masks. *Nature communications*, 12(1):1–6, 2021. 2

[11] Barmak Heshmat, Guy Satat, Christopher Barsi, and Ramesh Raskar. Single-shot ultrafast imaging using parallax-free alignment with a tilted lenslet array. In *CLEO: Science and Innovations*, pages STu3E–7. Optica Publishing Group, 2014. 3

[12] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *2011 International Conference on Computer Vision*, pages 287–294. IEEE, 2011. 2, 3, 9

[13] Jason Holloway, Aswin C Sankaranarayanan, Ashok Veeraraghavan, and Salil Tambe. Flutter shutter video camera for compressive sensing of videos. In *2012 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2012. 2, 3, 9

[14] Michael Iliadis, Leonidas Spinoulas, and Aggelos K Katsaggelos. Deepbinarymask: Learning a binary mask for video compressive sensing. *Digital Signal Processing*, 96:102591, 2020. 2, 3, 9

[15] Ulugbek S Kamilov. A parallel proximal algorithm for anisotropic total variation minimization. *IEEE Transactions on Image Processing*, 26(2):539–548, 2016. 7

[16] Roman Koller, Lukas Schmid, Nathan Matsuda, Thomas Niederberger, Leonidas Spinoulas, Oliver Cossairt, Guido Schuster, and Aggelos K Katsaggelos. High spatio-temporal resolution video with compressed sensing. *Optics express*, 23(12):15992–16007, 2015. 2, 3

[17] Alex Liberzon, Theo Käufer, Andreas Bauer, Peter Vennemann, and Erich Zimmer. Openpiv/openpiv-python: Openpiv-python v0.23.4, Jan. 2021. 8

[18] Rui Lima, Shigeo Wada, Shuji Tanaka, Motohiro Takeda, Takuji Ishikawa, Ken-ichi Tsubota, Yohsuke Imai, and Takami Yamaguchi. In vitro blood flow in a rectangular pdms microchannel: experimental observations using a confocal micro-piv system. *Biomedical microdevices*, 10:153–167, 2008. 2

[19] Yang Liu, Xin Yuan, Jinli Suo, David J Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2990–3006, 2018. 2, 3

[20] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J Brady. Coded aperture compressive temporal imaging. *Optics express*, 21(9):10526–10545, 2013. 2, 3

[21] LM Lourenco, A Krothapalli, and CA Smith. Particle image velocimetry. In *Advances in fluid mechanics measurements*, pages 127–199. Springer, 1989. 8

[22] Julien NP Martel, Lorenz K Mueller, Stephen J Carey, Piotr Dudek, and Gordon Wetzstein. Neural sensors: Learning pixel exposures for hdr imaging and video compressive sensing with programmable sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1642–1653, 2020. 2, 3, 9

[23] Amir Matin and Xu Wang. Compressive coded rotating mirror camera for high-speed imaging. In *Photonics*, volume 8, page 34. MDPI, 2021. 2

[24] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. Mc3d: Motion contrast 3d scanning. In *2015 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2015. 2

[25] Albert Abraham Michelson. *Experimental determination of the velocity of light: made at the US Naval Academy, Annapolis*, volume 1. US Nautical Almanac Office, 1880. 2

[26] Kristina Monakhova, Vi Tran, Grace Kuo, and Laura Waller. Untrained networks for compressive lensless photography. *Optics Express*, 29(13):20913–20929, 2021. 2, 3

[27] Manasi Muglikar, Leonard Bauersfeld, Diederik Paul Moeys, and Davide Scaramuzza. Event-based shape from polarization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1547–1556, 2023. 2

[28] SHR Müller, B Böhm, M Gleißner, R Grzeszik, S Arndt, and A Dreizler. Flow field measurements in an optically accessible, direct-injection spray-guided internal combustion engine using high-speed piv. *Experiments in fluids*, 48:281–290, 2010. 2

[29] Gergely Nagymáté and Rita M Kiss. Application of optitrack motion capture systems in human movement analysis: A systematic literature review. *Recent Innovations in Mechatronics*, 5(1.):1–9, 2018. 2

[30] Cindy M Nguyen, Julien NP Martel, and Gordon Wetzstein. Learning spatially varying pixel exposures for motion deblurring. *arXiv preprint arXiv:2204.07267*, 2022. 3, 9

[31] Basilio Pueo Ortega and José M Jiménez Olmedo. Application of motion capture technology for sport performance analysis. *Retos: nuevas tendencias en educación física, deporte y recreación*, (32):241–247, 2017. 2

[32] Travis Portz, Li Zhang, and Hongrui Jiang. Random coded sampling for high-speed hdr video. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2013. 2, 3

[33] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *Acm Siggraph 2006 Papers*, pages 795–804. 2006. 3

[34] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. In *CVPR 2011*, pages 329–336. IEEE, 2011. 2, 3, 9

[35] Aswin C Sankaranarayanan, Pavan K Turaga, Rama Chellappa, and Richard G Baraniuk. Compressive acquisition of linear dynamical systems. *SIAM Journal on Imaging Sciences*, 6(4):2109–2133, 2013. 2, 3

[36] Fulvio Scarano. Tomographic piv: principles and practice. *Measurement Science and Technology*, 24(1):012001, 2012. 2

[37] Yoav Y Schechner, Shree K Nayar, and Peter N Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on pattern analysis and machine intelligence*, 29(8):1339–1354, 2007. 4

[38] Shubham Sharma, Shubhankar Verma, Mohit Kumar, and Lavanya Sharma. Use of motion capture in 3d animation: motion capture systems, challenges, and recent trends. In *2019 international conference on machine learning, big data, cloud and parallel computing (comitcon)*, pages 289–294. IEEE, 2019. 2

[39] Mark Sheinin, Matthew O'Toole, and Srinivasa G Narasimhan. Deconvolving diffraction for fast imaging of sparse scenes. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2021. 2, 3, 4, 5, 8, 9

[40] Mark Sheinin, Dinesh N Reddy, Matthew O'Toole, and Srinivasa G Narasimhan. Diffraction line imaging. In *European Conference on Computer Vision*, pages 1–16. Springer, 2020. 2, 3, 4, 8

[41] Mark Sheinin, Yoav Y Schechner, and Kiriakos N Kutulakos. Computational imaging on the electric grid. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6437–6446, 2017. 2

[42] H Shen, L Gan, N Newman, Y Dong, C Li, Y Huang, and YC Shen. Spinning disk for compressive imaging. *Optics letters*, 37(1):46–48, 2012. 2

[43] H Shen, N Newman, L Gan, SC Zhong, Y Huang, and YC Shen. Compressed terahertz imaging system using a spin

disk. In *35th International Conference on infrared, millimeter, and terahertz waves*, pages 1–2. IEEE, 2010. 2

[44] Yaoyao Shi, Youwen Liu, Wei Sheng, Jiming Wang, and Tong Wu. Speckle rotation decorrelation based single-shot video through scattering media. *Optics express*, 27(10):14567–14576, 2019. 2

[45] Ho Shiraga, M Heya, A Fujishima, O Maegawa, K Shimada, Y Kato, T Yamanaka, and S Nakai. Laser-imploded core structure observed by using two-dimensional x-ray imaging with 10-ps temporal resolution. *Review of scientific instruments*, 66(1):722–724, 1995. 3

[46] Masada Tzabari and Yoav Y Schechner. Polarized optical-flow gyroscope. In *Proceedings of the European Conference on Computer Vision*, pages 363–381. Springer, 2020. 2

[47] BW Van Oudheusden. Piv-based pressure measurement. *Measurement Science and Technology*, 24(3):032001, 2013. 2

[48] Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):671–686, 2010. 2, 3

[49] Zihao Wang, Leonidas Spinoulas, Kuan He, Lei Tian, Oliver Cossairt, Aggelos K Katsaggelos, and Huaijin Chen. Compressive holographic video. *Optics express*, 25(1):250–262, 2017. 2, 3

[50] Gil Weinberg and Ori Katz. 100,000 frames-per-second compressive imaging with a conventional rolling-shutter camera by random point-spread-function engineering. *Optics Express*, 28(21):30616–30625, 2020. 2, 3, 4, 8

[51] Charles Wheatstone. An account of some experiments to measure the velocity of electricity and the duration of electric light. *Philosophical Transactions of the Royal Society of London*, (124):583–591, 1834. 2

[52] Jinhui Xiong, Ramzi Idoughi, Andres A Aguirre-Pablo, Abdulrahman B Aljedaani, Xiong Dun, Qiang Fu, Sigurdur T Thoroddsen, and Wolfgang Heidrich. Rainbow particle imaging velocimetry for dense 3d fluid velocity imaging. *ACM Transactions on Graphics (TOG)*, 36(4):1–14, 2017. 8