

Coded Strobing Photography: Compressive Sensing of High Speed Periodic Videos

Ashok Veeraraghavan*, *Member, IEEE*, Dikpal Reddy*, *Student Member, IEEE*,
and Ramesh Raskar, *Member, IEEE*

Abstract—We show that, via temporal modulation, one can observe and capture a high-speed periodic video well beyond the abilities of a low-frame-rate camera. By strobing the exposure with unique sequences within the integration time of each frame, we take coded projections of dynamic events. From a sequence of such frames, we reconstruct a high-speed video of the high-frequency periodic process. Strobing is used in entertainment, medical imaging, and industrial inspection to generate lower beat frequencies. But this is limited to scenes with a detectable single dominant frequency and requires high-intensity lighting. In this paper, we address the problem of sub-Nyquist sampling of periodic signals and show designs to capture and reconstruct such signals. The key result is that for such signals, the Nyquist rate constraint can be imposed on the strobe rather than the sensor rate. The technique is based on intentional aliasing of the frequency components of the periodic signal while the reconstruction algorithm exploits recent advances in sparse representations and compressive sensing. We exploit the sparsity of periodic signals in the Fourier domain to develop reconstruction algorithms that are inspired by compressive sensing.

Index Terms—Computational imaging, high speed imaging, compressive sensing, compressive video sensing, stroboscopy.

I. INTRODUCTION

Periodic signals are all around us. Several human and animal biological processes such as heart-beat, breathing, several cellular processes, industrial automation processes and everyday objects such as hand-mixer and blender all generate periodic processes. Nevertheless, we are mostly unaware of the inner workings of some of these high-speed processes because they occur at a far greater speed than can be perceived by the human eye. Here, we show a simple but effective technique that can turn an off-the-shelf video camera into a powerful high-speed video camera for observing periodic events.

Strobing is often used in entertainment, medical imaging and industrial applications to visualize and capture high-speed visual phenomena. Active strobing involves illuminating the scene with a rapid sequence of flashes within a frame time. The classic example is Edgerton's Rapatron to capture a golf swing [13]. In modern sensors, it is achieved passively by multiple-exposures within a frame time [36][28] or fluttering [29]. We use the term 'strobing' to indicate both active illumination and passive sensor methods.

In case of periodic phenomenon, strobing is commonly used to achieve aliasing and generate lower beat frequencies. While strobing performs effectively when the scene consists of a single frequency with a narrow sideband, it is difficult to visualize multiple or a wider band of frequencies simultaneously.

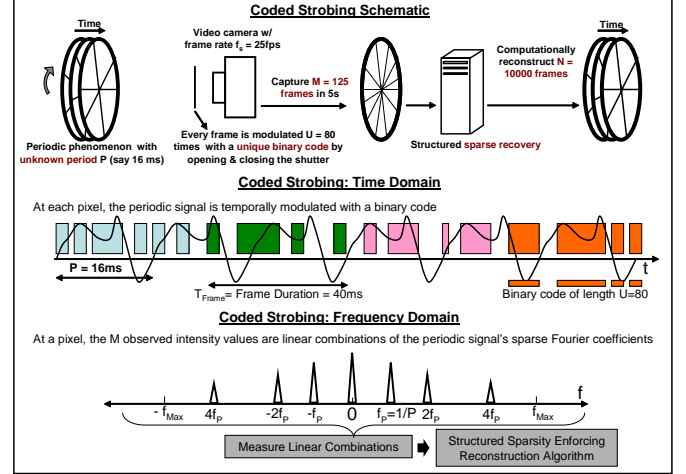


Fig. 1: CSC: A fast periodic visual phenomenon is recorded by a normal video camera (25 fps) by randomly opening and closing the shutter at high speed (2000 Hz). The phenomenon is accurately reconstructed from the captured frames at the high-speed shutter rate (2000 fps).

Instead of direct observation of beat frequencies, we exploit a computational camera approach based on different sampling sequences. The key idea is to measure appropriate linear combinations of the periodic signal and then decode the signal by exploiting the sparsity of the signal in Fourier domain. We observe that by coding during the exposure duration of a low-frame-rate (e.g., 25 fps) video camera, we can take appropriate projections of the signal needed to reconstruct a high-frame-rate (e.g., 2000 fps) video. During each frame, we strobe and capture a coded projection of the dynamic event and store the integrated frame. After capturing several frames, we computationally recover the signal independently at each pixel by exploiting the Fourier sparsity of periodic signals. Our method of coded exposure for sampling periodic signals is termed 'coded strobing' and we call our camera the 'coded strobing camera' (CSC). Figure 1 illustrates the functioning of CSC.

A. Contributions

- We show that sub-Nyquist sampling of periodic visual signals is possible and that such signals can be captured and recovered using a coded strobing computational camera.
- We develop a sparsity-exploiting reconstruction algorithm and expose connections to Compressive Sensing.
- We show that the primary benefit of our approach over traditional strobing is, increased light-throughput and

* Ashok Veeraraghavan and Dikpal Reddy contributed equally to this work.

ability to tackle multiple frequencies simultaneously post-capture.

B. Benefits and limitations

The main constraint for recording a high-speed event is **light throughput**. We overcome this constraint for periodic signals via sufficient exposure duration (in each frame) and extended observation window (multiple frames). For well-lit non-periodic events, high-speed cameras are ideal. For a static snapshot, a short exposure photo (or single frame of the high-speed camera) is sufficient. In both cases, light throughput is limited but unavoidable. Periodic signals can also be captured with high-speed camera. But one will need a well-lit scene or must illuminate it with unrealistic bright lights. For example, if we use a 2000 fps camera for vocal cord analysis instead of strobing using a laryngoscope, we would need a significantly brighter illumination source and this creates the risk of burn injuries to the throat. A safer option would be 25 fps camera with strobed light source and then exploit the periodicity of vocal fold movement. Here, we show that an even better option in terms of light-throughput is a computational camera approach. Further, the need to know frequency of the signal at capture-time is also avoided. Moreover, the computational recovery algorithm can tackle the presence of multiple fundamental frequencies in a scene, which poses a challenge to traditional strobing.

C. Related work

High-speed imaging hardware: Capturing high-speed events with fast, high-frame rate cameras require imagers with high photoresponsivity at short integration times, synchronous exposure and high-speed parallel readout due to the necessary bandwidth. In addition, they suffer from challenging storage problems. A high-speed camera also fails to exploit the inter-frame coherence, while our technique takes advantage of a simplified model of motion. Edgerton and others have shown visually stunning results for high-speed objects using extremely narrow-duration flash [13]. These snapshots capture an instant of the action but fail to indicate the general movement in the scene. Multiple low-frame rate cameras can be combined to create high-speed sensing. Using a staggered exposure approach, Shechtman et al. [33] used frames captured by multiple co-located cameras with overlapped exposure time. This staggered exposure approach also assisted a novel reconfigurable multi-camera array [37]. Although there are very few methods to super-resolve a video temporally [15], numerous super-resolution techniques have been proposed to increase the spatial resolution of images. In [17], super-resolution technique to reconstruct a high-resolution image from a sequence of low-resolution images was proposed by backprojection method. A method to do super-resolution on a low quality image of a moving object by first tracking it, estimating motion and deblurring the motion blur and creating a high quality image was proposed in [4]. Freeman et al. [14] proposed a learning based technique for superresolution from one image where the high frequency components like edges of an image are filled by patches obtained from examples with similar low resolution properties. Finally, fundamental limits

on super-resolution for reconstruction based algorithms have been explored in [1][22].

Stroboscopy and periodic motion: Stroboscopes (from the Greek word $\sigma\tau\rho\omega\beta\omega\sigma$ for ‘whirling’) play an important role in scientific research, to study machinery in motion, in entertainment and medical imaging. Muybridge in his pioneering work used multiple triggered cameras to capture high-speed motion of animals [25] and proved that all four of a horse’s hooves left the ground at the same time during a gallop. Edgerton also used flashing lamp to study machine parts in motion [13]. The most common approaches for “freezing” or “slowing down” the movement are based on temporal aliasing. In medicine, stroboscopes are used to view the vocal cords for diagnosis. The patient hums or speaks into a microphone which in turn activates the stroboscope at either the same or a slightly lower frequency [20],[30]. However, in all healthy humans, vocal-fold vibrations are aperiodic to a greater or lesser degree. Therefore, strobolaryngoscopy does not capture the fine detail of each individual vibratory cycle; rather, it shows a pattern averaged over many successive nonidentical cycles [24][32]. Modern stroboscopes for machine inspection [11] are designed for observing fast repeated motions and for determining RPM. The idea can also be used to improve spatial resolution by introducing high-frequency illumination [16].

Processing: In computer vision, periodic motion of humans has received significant attention. Seitz et al. [31] introduced a novel motion representation, called the period trace, that provides a complete description of temporal variations in a cyclic motion, which can be used to detect motion trends and irregularities. A technique to repair videos with large static background or cyclic motion was presented in [18]. Laptev et al. [19] presented a method to detect and segment periodic motion based on sequence alignment without the need for camera stabilization and tracking. [5] exploited periodicity of moving objects to perform 3D reconstruction by treating frames with same phase to be of same pose observed from different views. In [34], the authors showed a strobe based approach for capturing high-speed motion using multiexposure images obtained within a single frame of a camera. The images of a baseball appear as distinct non-overlapping positions in the image. High temporal and spatial resolution can be obtained via a hybrid imaging device which consists of a high spatial resolution digital camera in conjunction with a high frame-rate but low resolution video camera [6]. In cases where the motion can be modeled as linear, there have been several interesting methods to engineer the motion blur point spread function so that the blur induced by the imaging device is invertible. These include coding the exposure [29] and moving the sensor during the exposure duration [21]. The method presented in this paper tackles a different but broadly related problem of reconstructing periodic signals from very low-speed images acquired via a conventional video camera (albeit enhanced with coded exposure).

Comparison with flutter shutter: In [29], the authors showed that by opening and closing the shutter according to an optimized coded pattern during the exposure duration of a

photograph, one can preserve high-frequency spatial details in the blurred captured image. The image can be then de-blurred using a manually specified point-spread function. Similarly, we open and close the shutter according to a coded pattern and this code is optimized for capture. Nevertheless, there are significant differences in motion models and reconstruction procedures of both these methods. In flutter shutter (FS), a constant velocity linear motion model was assumed and de-blurring was done in blurred pixels along the motion direction. On the other hand, CSC works even on very complicated motion models as long as the motion is periodic. In CSC each of the captured frames is the result of modulation with a different binary sequence whereas in FS a single frame is modulated with a ‘all-pass’ code. Further, our method contrasts fundamentally with FS in reconstruction of the frames. In FS the system of equations is not under-determined whereas in CSC we have a severely under-determined system. We overcome this problem by ℓ_1 -norm regularization, appropriate for enforcing sparsity of periodic motion in time. In FS a single system of equations is solved for entire image whereas in CSC at each pixel we temporally reconstruct the periodic signal by solving an under-determined system.

D. Capture and reconstruction procedure

The sequence of steps involved in capture and reconstruction of a high-speed periodic phenomenon with typical physical values are listed below with references to appropriate sections for detailed discussion.

- Goal: Using a 25 fps camera and a shutter which can open and close at 2000 Hz, capture a high-speed periodic phenomenon of unknown period by observing for 5s.
- The length of the binary code needed is $N = 2000 \times 5 = 10000$. For an upsampling factor of $U = 2000/25 = 80$, find the optimal pseudo random code of length N (Section III-A).
- Capture $M = 25 \times 5 = 125$ frames by fluttering the shutter according to the optimal code. Each captured frame is an integration of the incoming visual signal modulated with a corresponding subsequence of binary values of length $U = 80$ (Section II-C).
- Estimate the fundamental frequency of the periodic signal (Section II-D3).
- Using the estimated fundamental frequency, at each pixel reconstruct the periodic signal of length $N = 10000$ from $M = 125$ values by recovering the signal’s sparse Fourier coefficients (Section II-D).

II. STROBING AND LIGHT MODULATION

A. Traditional sampling techniques

Sampling is the process of converting a continuous domain signal into a set of discrete samples in a manner that allows approximate or exact reconstruction of the continuous domain signal from just the discrete samples. The most fundamental result in sampling is that of Nyquist-Shannon sampling theorem. Figure 2 provides a graphical illustration of traditional sampling techniques applied to periodic signals.

Nyquist sampling: Nyquist-Shannon sampling states that when a continuous domain signal is band-limited to $[0, f_0]$

Hz, one can exactly reconstruct the band-limited signal, by just observing discrete samples of the signal at a sampling rate f_s greater than $2f_0$ [27]. When the signal has frequency components that are higher than the prescribed band-limit, then during the reconstruction, the higher frequencies get aliased as lower frequencies making the reconstruction erroneous (see Figure 2(Right)(c)). If the goal is to capture a signal whose maximum frequency f_{Max} is 1000 Hz, then one needs a high-speed camera capable of 2000 fps in order to acquire the signal. Such high-speed video cameras are light limited and expensive.

Band-pass sampling (strobging): If the signal is periodic as shown in Figure 2(Left)(a), then we can intentionally alias the periodic signal by sampling at a frequency very close to the fundamental frequency of the signal as shown in Figure 2(Left)(e). This intentional aliasing allows us to measure the periodic signal. This technique is commonly used for vocal fold visualization [24][32]. However, traditional strobing suffers from the following limitations. The frequency of the original signal must be known at capture-time so that one may perform strobing at the right frequency. Secondly, the strobe signal must be ‘ON’ for a very short duration so that the observed high-speed signal is not smoothed out and this makes traditional strobing light-inefficient. Despite this handicap, traditional strobing is an extremely interesting and useful visualization tool (and has found several applications in varying fields).

Non-uniform sampling: With periodic sampling, aliasing occurs when the sampling rate is not adequate because, all frequencies of the form $f_1 + k \cdot f_s$ (k an integer) lead to identical samples. One method to counter this problem is to employ non-uniform or random sampling [7][23]. The key idea in non-uniform sampling [7][23] is to ensure a set of sampling instants such that the observation sequence for any two frequencies are different at least in one sampling instant. This scheme has never found widespread practical applicability because of its noise sensitivity and light inefficiency.

B. Periodic signals

Since, the focus of this paper is on high-speed video capture of periodic signals, we first study the properties of such signals.

1) *Fourier domain properties of periodic signals:* Consider a signal $x(t)$, which has a period $P = 1/f_P$ and a bandlimit f_{Max} . Since the signal is periodic, we can express it as,

$$x(t) = x_{DC} + \sum_{j=1}^{j=Q} a_j \cos(2\pi j f_P t) + b_j \sin(2\pi j f_P t) \quad (1)$$

Therefore, the Fourier transform of the signal $x(t)$ contains energy only in the frequencies corresponding to $j f_P$, where $j \in \{-Q, -(Q-1), \dots, 0, 1, \dots, Q\}$. Thus, a periodic signal has a maximum of $(K = 2Q + 1)$ non-zero Fourier coefficients. Therefore, periodic signals by definition, have a very sparse representation in the Fourier domain. Recent advances in the field of compressed sensing (CS) [12][9][2][8][35] have developed reliable recovery algorithms for inferring sparse representations if one can measure arbitrary linear combinations

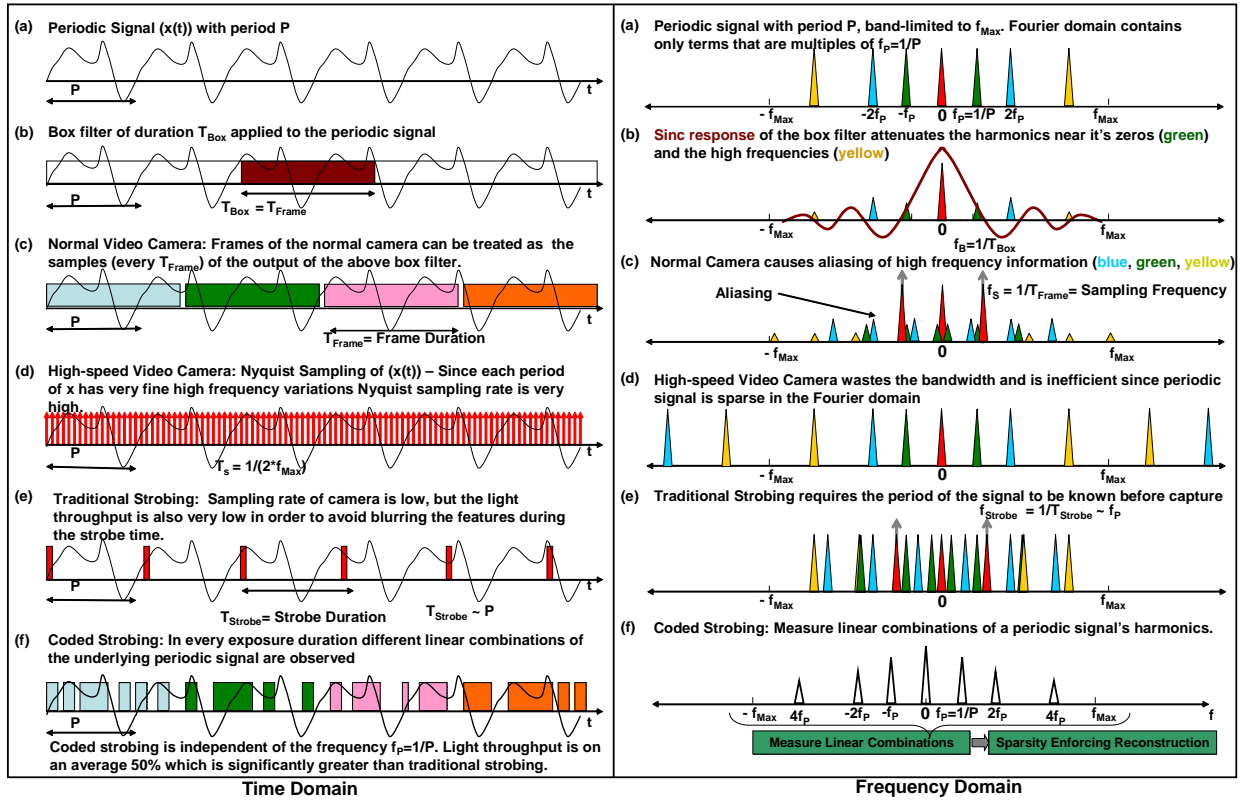


Fig. 2: Time domain (Left) and the corresponding frequency domain (Right) characteristics of various sampling techniques as applicable to periodic signals. Note that capturing high-speed visual signals using normal camera can result in attenuation of high frequencies ((b) and (c)) whereas a high-speed camera demands large bandwidth (d) and traditional strobing is light-inefficient (e). Coded strobing is shown in (f). To illustrate sampling only two replicas have been shown and note that colors used in time domain and frequency domain are unrelated.

of the signals. Here, we propose and describe a method for measuring such linear combinations and use the reconstruction algorithms inspired by CS to recover the underlying periodic signal from its low-frame-rate observations.

2) Effect of visual texture on periodic motion: Visual texture on surfaces exhibiting periodic motion introduces high frequency variations in the observed signal (Figure 3(d)). As a very simple instructive example consider the fan shown in Figure 3(a). The fan rotates at a relatively slow rate of 8.33 Hz. This would seem to indicate that in order to capture the spinning fan one only needs a 16.66 fps camera. During exposure time of 60 ms of a 16.66 Hz camera, the figure ‘1’ written on the fan blade completes about half a revolution blurring it out (Figure 3(b)). Shown in Figure 3(c) is the time profile of the intensity of a single pixel using a high-speed video camera. Note that the sudden drop in intensity due to the dark number ‘1’ appearing on the blades persists only for about 1 millisecond. Therefore, we need a 1000 fps high-speed camera to observe the ‘1’ without any blur. In short, the highest temporal frequency observed at a pixel is a product of the highest frequency of the periodic event in time and the highest frequency of the spatial pattern on the objects across the direction of motion. This makes the capture of high-speed periodic signals with texture more challenging.

3) Quasi-periodic signals: Most real world “periodic signals” are not exactly so, but almost; there are small changes in the period of the signal over time. We refer to such broader class

of signals as quasi-periodic. For example, the Crest toothbrush we use in our experiments exhibits a quasi-periodic motion with fundamental frequency that varies between 63 – 64 Hz. Figure 4(a) shows few periods of a quasi-periodic signal at a pixel of a vibrating tooth brush. Variation in fundamental frequency f_P , between 63 and 64 Hz, over time can be seen in (b). Variation in f_P of a quasi-periodic signal is reflected in its Fourier transform which contains energy not just at multiples jf_P but in small band around jf_P . Nevertheless, like periodic signals, the Fourier coefficients are concentrated at jf_P (Figure 4(c)) and are sparse in the frequency domain. The coefficients are distributed in a band $[jf_P - j\Delta f_P, jf_P + j\Delta f_P]$. For example, $\Delta f_P = 0.75$ Hz in Figure 4(d).

C. Coded exposure sampling (or Coded strobing)

The key idea is to measure appropriate linear combinations of the periodic signal and then recover the signal by exploiting the sparsity of the signal in Fourier domain (Figure 5). Observe that by coding the incoming signal during the exposure duration, we take appropriate projections of the desired signal.

1) Camera observation model: Consider a luminance signal $x(t)$. If the signal is band-limited to $[-f_{Max}, f_{Max}]$, then in order to accurately represent and recover the signal, we only need to measure samples of the signal that are $\delta t = 1/(2f_{Max})$ apart where δt represents the temporal resolution with which we wish to reconstruct the signal. If the total time of observing

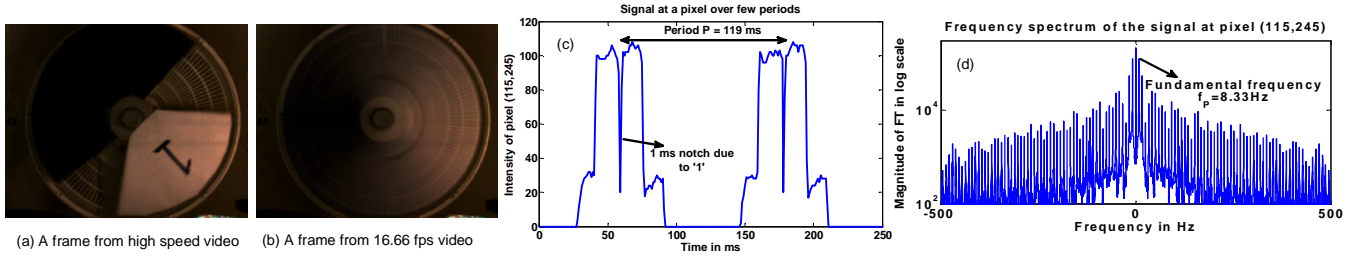


Fig. 3: (a) Video of a fan from a high-speed camera (b) A 16.66 Hz camera blurs out the '1' in the image (c) Few periods of the signal at a pixel where the figure '1' passes. Note the notch of duration 1 ms in the intensity profile. (d) Fourier transform of the signal in (c). Notice the higher frequency components in a signal with low fundamental frequency f_P .

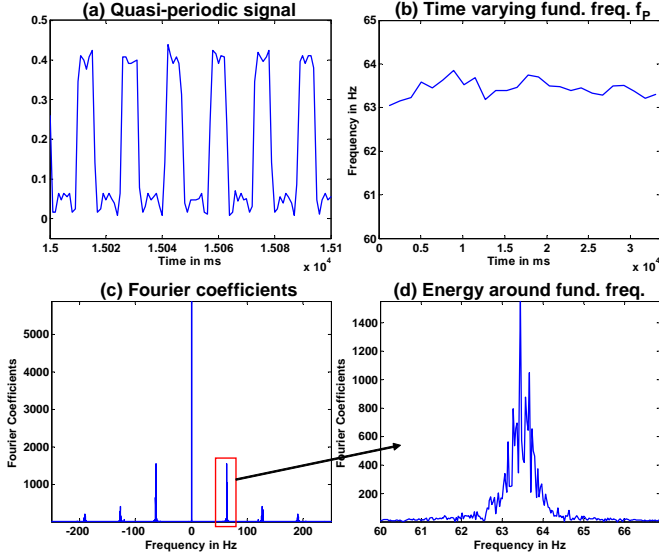


Fig. 4: (a) Six periods of a $N = 32768$ ms long quasi-periodic signal at a pixel of a scene captured by 1000 fps high-speed camera. (b) Fundamental frequency f_P varying with time. (c) Fourier coefficients of the quasi-periodic signal shown in (a). (d) On zoom we notice that the signal energy is concentrated in a band around the fundamental frequency f_P and its harmonics.

the signal is $N\delta t$, then the N samples can be represented in a N dimensional vector x .

In a normal camera, the radiance at a single pixel is integrated during the exposure time, and the sum is recorded as the observed intensity at a pixel. Instead of integrating during the entire frame duration, we perform amplitude modulation of the incoming radiance values, before integration. Then the observed intensity values y at a given pixel can be represented as

$$y = Cx + \eta, \quad (2)$$

where the $M \times N$ matrix C performs both the modulation and integration for frame duration, and η represents the observation noise. Figure 5 shows the structure of matrix C . If the camera observes a frame every T_s seconds, the total number of frames/observations would be $M = N\delta t/T_s$ and so y is a $M \times 1$ vector. The camera sampling time T_s is far larger than the time resolution we would like to achieve (δt), therefore $M \ll N$. The upsampling factor (or decimation ratio) of CSC can be defined as,

$$\text{Upsampling factor} = U = \frac{N}{M} = \frac{2f_{Max}}{f_s}. \quad (3)$$

For example, in the experiment shown in Figure 15, $f_{Max} = 1000$ Hz, and $f_s = 25$ fps. Therefore, the upsampling factor achieved is 80, i.e., the frame-rate of CSC is eighty times smaller than that of an equivalent high-speed video camera. Even though, the modulation function can be arbitrary, in practice it is usually restricted to be binary (open or close shutter). Effective modulation can be achieved with codes that have a 50% transmission, i.e., the shutter is open for 50% of the total time, thereby limiting light-loss at capture-time to just 50%.

2) *Signal model:* If x , the luminance at a pixel is bandlimited it can be represented as,

$$x = Bs, \quad (4)$$

where, the columns of B contain Fourier basis elements. Moreover, since the signal $x(t)$ is assumed to be periodic, we know that the basis coefficient vector s is sparse as shown in Figure 5. Putting together the signal and observation model, the intensities in the observed frames are related to the basis coefficients as,

$$y = Cx + \eta = CBs + \eta = As + \eta, \quad (5)$$

where A is the effective mixing matrix of the forward process. Recovery of the high-speed periodic motion x amounts to solving the linear system of equations (5).

D. Reconstruction algorithms

To reconstruct the high-speed periodic signal x , it suffices to reconstruct its Fourier coefficients s from modulated intensity observations y of the scene.

Unknowns, measurements and sparsity: In (5), the number of unknowns exceeds the number of known variables by a factor U (typically 80) and hence the system of equations (5) is severely under-determined ($M \ll N$). To obtain robust solutions, further knowledge about the signal must be used. Since the Fourier coefficients s , of a periodic signal x , are sparse, a reconstruction technique enforcing sparsity of s could still hope to recover the periodic signal x .

We present two reconstruction algorithms, one which enforces the sparsity of the Fourier coefficients and is inspired by compressive sensing and other which additionally enforces the structure of the sparse Fourier coefficients.

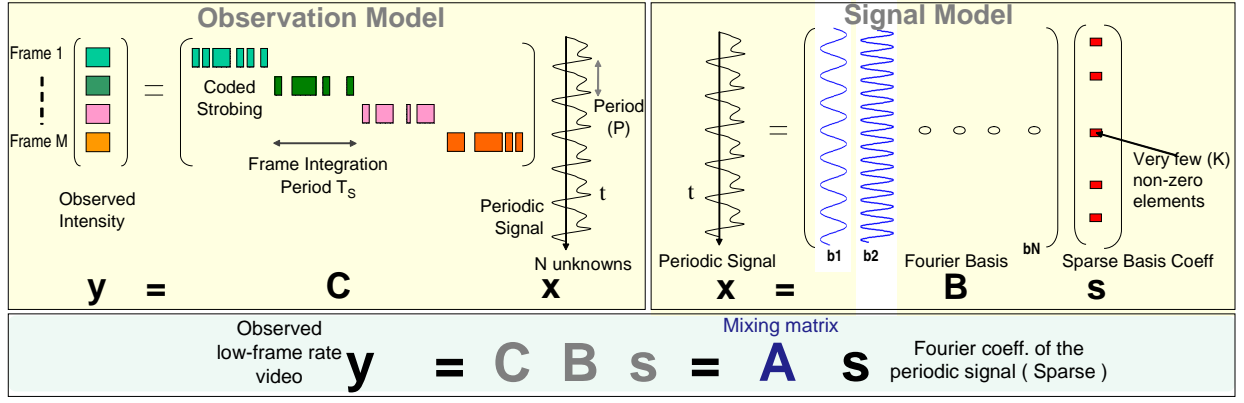


Fig. 5: Observation model shows the capture process of the CSC where different colors correspond to different frames and the binary shutter sequence is depicted using the presence or absence of color. Note that each frame uses a different binary sub-sequence. The signal model illustrates the sparsity in the frequency spectrum of a periodic signal.

1) *Sparsity enforcing reconstruction:* Estimating a sparse vector s (with K non-zero entries) that satisfies $y = As + \eta$, can be formulated as an ℓ_0 optimization problem:

$$(P0): \quad \min \|s\|_0 \quad \text{s.t.} \quad \|y - As\|_2 \leq \epsilon. \quad (6)$$

Although for general s this is a NP-hard problem, for K sufficiently small the equivalence between ℓ_0 and ℓ_1 -norm [8] allows us to reformulate the problem as one of ℓ_1 -norm minimization, which is a convex program with very efficient algorithms [12][8][2].

$$(P1): \quad \min \|s\|_1 \quad \text{s.t.} \quad \|y - As\|_2 \leq \epsilon \quad (7)$$

The parameter ϵ allows for the variation in the modeling of signal's sparsity and/or noise in the observed frames. In practice, it is set to a fraction of captured signal energy (e.g., $\epsilon = 0.03\|y\|_2$) and is dictated by the prior knowledge about camera noise in general and the extent of periodicity of the captured phenomenon. An interior point implementation (BPDN) of (P1) is used to accurately solve for s . Instead, in most experiments in this paper, at the cost of minor degradation in performance we use CoSaMP [26], a faster greedy algorithm to solve (P0). Both (P0) and (P1) don't take into account the structure in the sparse coefficients of the periodic signal. By additionally enforcing structure of the sparse coefficients s , we achieve robustness in recovery of the periodic signal.

2) *Structured sparse reconstruction:* We recall that periodic/quasi-periodic signals are (a) sparse in the Fourier basis and (b) if the period is $P = 1/f_P$, the only frequency content the signal has is in the small bands at the harmonics jf_P , j an integer. Often, the period P is not known a priori. If the period is known or can be estimated from the data y , then for a hypothesized fundamental frequency f_H , we can construct a set S_{f_H} with basis elements $[jf_H - \Delta f_H, jf_H + \Delta f_H]$, for $j \in \{-Q, \dots, 0, 1, \dots, Q\}$ such that all the sparse Fourier coefficients will lie in this smaller set. Now the problem (P0) can instead be reformulated as

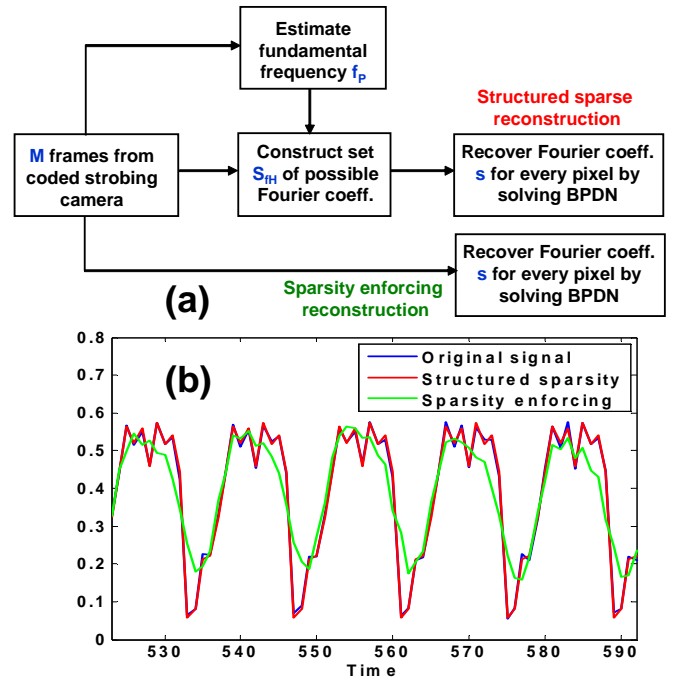


Fig. 6: (a) Overview of structured sparse and sparsity enforcing reconstruction algorithms (b) Five periods of a noisy (SNR=35 dB) periodic signal x ($P = 14$ units). Signal recovered by structured and normal sparsity enforcing reconstruction are also shown.

$$(P_{Structured}): \quad \min \|s\|_0 \quad \text{s.t.} \quad (8) \\ \|y - As\|_2 \leq \epsilon \quad \text{and} \\ \text{nonZero}(s) \in S_{f_H} \quad \text{for some } f_H \in [0, f_{Max}].$$

where $\text{nonZero}(s)$ is a set containing all the non-zero elements in the reconstructed s . Since the extent of quasi-periodicity is not known a priori, the band Δf_H is chosen safely large and the non-zero coefficients continue to remain sparse in the set S_{f_H} . Intuitively, problem $P_{Structured}$ gives a better sparse solution compared to (P0) since the non-zero coefficients are searched over a smaller set S_{f_H} . An example of a periodic signal and its recovery using sparsity enforcing (P1) and structured sparsity are shown in Figure 6(b). The

recovery using $P_{Structured}$ is exact whereas (P0) fails to recover the high-frequency components.

The restatement of the problem provides two significant advantages. Firstly, it reduces the problem search space of the original ℓ_0 formulation. To solve the original ℓ_0 formulation, one has to search over $^N C_K$ sets. For example, if we observe a signal for 5 seconds at 1 ms resolution, then N is 5000 and $^N C_K$ is prohibitively large (10^{212} for $K = P = 100$). Secondly, this formulation implicitly enforces the quasi-periodicity of the recovered signal and this extra constraint allows us to solve for the unknown quasi-periodic signal with far fewer measurements than would otherwise be possible. The type of algorithms which exploit further statistical structure in the support of the sparse coefficients come under model-based compressive sensing [3].

3) *Knowledge of fundamental frequency: Structured sparse reconstruction* performs better over a larger range of upsampling factors and since the structure of non-zero coefficients is dependent on fundamental frequency f_P , we estimate it first.

Identification of fundamental frequency: For both periodic and quasi-periodic signals we solve a sequence of least-square problems to identify the fundamental frequency f_P . For a hypothesized fundamental frequency f_H , we build a set S_{f_H} with only the frequencies $j f_H$ (for both periodic and quasi-periodic signals). Truncated matrix A_{f_H} is constructed by retaining only the columns with indices in S_{f_H} . Non-zero coefficients \hat{s}_{f_H} are then estimated by solving the equation $y = A_{f_H} s_{f_H}$ in a least-squares sense. We are interested in f_H which has a small reconstruction error $\|y - \hat{y}_{f_H}\|$ (or largest output SNR) where $\hat{y}_{f_H} = A_{f_H} \hat{s}_{f_H}$. If f_P is the fundamental frequency, then all the sets S_{f_H} , where f_H is a factor of f_P , will provide a good fit to the observed signal y . Hence, the plot of output SNR has multiple peaks corresponding to the good fits. From these peaks we pick the one with largest f_H . In Figure 7, we show results of experiments on synthetic datasets, under two scenarios: noisy signal and quasi-periodicity. We note that even when (a) the signal is noisy and (b) when the quasi-periodicity of the signal increases, the last peak in the SNR plot occurs at fundamental frequency f_P . We generate quasi-periodic signals from periodic signals by warping the time variable. Note that, solving a least squares problem for a hypothesized fundamental frequency f_H is equivalent to solving $P_{structured}$ with $\Delta f_H = 0$. Setting $\Delta f_H = 0$ eases the process of finding the fundamental frequency by avoiding the need to set the parameter Δf_H appropriate for both the captured signal and f_H . This is especially useful for quasi-periodic signals where a priori knowledge of quasi-periodicity is not available.

III. DESIGN ANALYSIS

In this section, we analyze important design issues and gain a better understanding of the performance of coded strobing method through experiments on synthetic examples.

A. Optimal code for coded strobing

Theoretically optimal code: The optimization problems (6) and (7) give unique and exact solutions provided the under-determined matrix A satisfies the *restricted isometry property*

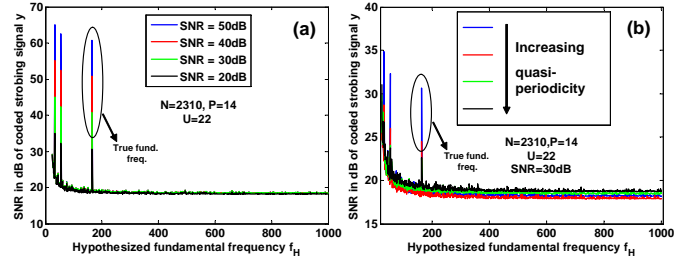


Fig. 7: Identifying the fundamental frequency f_P . Output SNR $\|y\|/\|y - \hat{y}_{f_H}\|$ in dB is plotted against hypothesized fundamental frequency f_H . (a) Plot of SNR as the noise in y is varied. Note that the last peak occurs at $f_H = 165$ ($= \frac{N}{P}$). (b) Plot of SNR with varying level of quasi-periodicity.

(RIP) [10]. Since the location of the K non-zeros of the sparse vector s which generates the observation y is not known a priori, RIP demands that all sub-matrices of A with $2K$ columns have a low condition number. In other words, every possible restriction of $2K$ columns are nearly orthonormal and hence isometric. Evaluating RIP for a matrix is a combinatorial problem since it involves checking the condition number of all $^N C_{2K}$ submatrices.

Alternately, matrix A satisfies RIP if every row of C is incoherent with every column of B . In other words, no row of C can be sparsely represented by columns of B . Tropp et al. [35] showed in a general setting that if the code matrix C is drawn from a IID Rademacher distribution, the resulting mixing matrix A satisfies RIP with a high probability. It must be noted that a modulation matrix C with entries ‘+1’, ‘-1’ is implementable but would involve using a beam splitter and two cameras in place of one. Due to ease of implementation (details in section IV), for modulation we use a binary ‘1’, ‘0’ code matrix C as described in section II-C1. For a given signal length N and an upsampling factor U we would like to pick a binary ‘1’, ‘0’ code which results in mixing matrix A , optimal in the sense of RIP.

Note that the sparsity of quasi-periodic signals is structured and the non-zero elements occur at regular intervals. Hence, unlike the general setting, RIP should be satisfied and evaluated over only a select subset of columns. Since the fundamental frequency f_P of the signal is not known a priori, it suffices if the isometry is evaluated over a sequence of matrices \bar{A} corresponding to hypothesized fundamental frequency f_H . Hence, for a given N and U , a code matrix C which results in smallest condition number over all the sequence of matrices \bar{A} is desired. In practice, such a C is sub-optimally found by randomly generating the binary codes tens of thousand times and picking the best one.

Compared to a normal camera, CSC blocks half the light but captures all the frequency content of the periodic signal. The sinc response of the box filter of a normal camera attenuates the harmonics near its zeros as well as the higher frequencies as shown in Figure 2(b). To avoid the attenuation of harmonics, the frame duration of the camera has to be changed appropriately. But, this is undesirable since most cameras come with a discrete set of frame rates. Moreover, it is hard to have a priori knowledge of the signal’s period. This

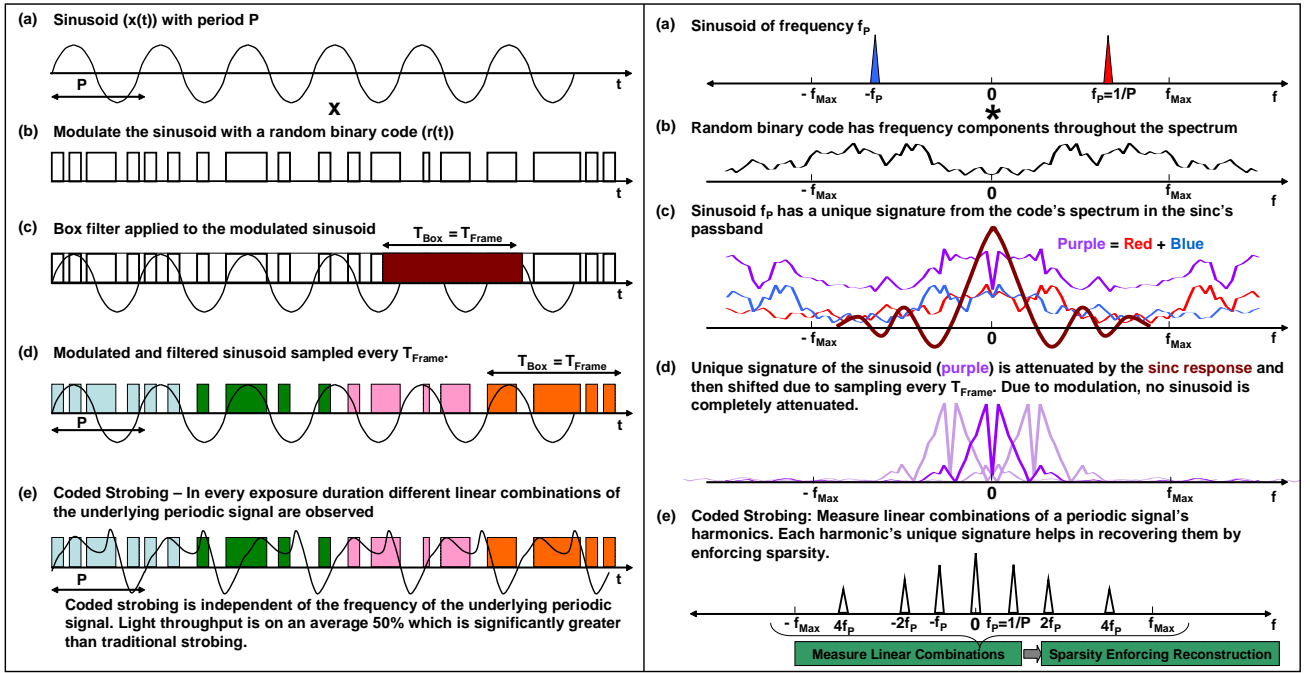


Fig. 8: Time domain (Left) and corresponding frequency domain (Right) understanding of CSC. Shown in (a) is a single sinusoid. (b),(c) & (d) show the effect of coded strobing capture on the sinusoid. (e) coded strobing capture of multiple sinusoids is simply a linear combination of the sinusoids.

problem is entirely avoided by modulating the incoming signal with a pseudo-random binary sequence. Shown in Figure 8 is the temporal and frequency domain visualization of the effect of CSC on a single harmonic. Modulation with a pseudo-random binary code spreads the harmonic across the spectrum. Thus, every harmonic irrespective of its position avoids the attenuation, the sinc response causes.

We perform numerical experiments to show the effectiveness of CSC (binary code) over the normal camera (all ‘1’ code). Shown in Table I are the comparison of the largest and smallest condition numbers of the matrix \bar{A} arising in CSC and normal camera. For a given signal length $N = 5000$ and upsampling factor $U = 25$ (the second column in Table 1), we vary the period P and generate different matrices \bar{A} for both CSC and normal camera. The largest condition number (1.8×10^{19}) of mixing matrix \bar{A} of a normal camera occurs for signal of period $P = 75$. Similarly, the smallest condition number occurs for $P = 67$. On the other hand, the mixing matrix \bar{A} of CSC has significantly lower maximum (at $P = 9$) and minimum (at $P = 67$) condition numbers. Note that the largest and smallest condition number of CSC matrices \bar{A} across different upsampling factors U are significantly smaller compared to those of normal camera matrices. This indicates that when the period of the signal is not known a priori, it is prudent to use CSC over normal camera.

Performance evaluation: We perform simulations on periodic signals to compare the performance of *sparsity enforcing* and *structured sparse* reconstruction algorithms on CSC frames, *structured sparse* reconstruction on normal camera frames and traditional strobing. SNR plots of the reconstructed signal using the four approaches for varying period P , upsampling factor U and noise level in y are shown in Figure 9. The

signal length is fixed to $N = 2000$ units. The advantage of *structured sparse* reconstruction is apparent from comparing blue and red plots. The advantage of CSC over normal camera can be seen by comparing blue and black plots. Note that the normal camera performs poorly when the upsampling factor U is a multiple of the period P .

B. Experiments on a synthetic animation

We perform experiments on a synthetic animation of a fractal to show the efficacy of our approach. We also analyze the performance of the algorithm under various noisy scenarios. We assume that at every $\delta t = 1$ ms, a frame of the animation is being observed and that the animation is repetitive with $P = 25$ ms (25 distinct images in the fractal). Two such frames are shown in Figure 10(a). A normal camera running at $f_s = 25$ fps will integrate 40 frames of the animation into a single frame, resulting in blurred images. Two images from a 25 fps video are shown in (b). By performing amplitude modulation at the shutter, as described in II-C1, the CSC obtains frames at the same rate as that of the normal camera (25 fps) but with the images encoding the temporal movement occurring during the integration process of the camera sensor. Two frames from the CSC are shown in (c). Note that in images (b) & (c) and also images in other experiments we rescaled the intensities appropriately for better display. For our experiment, we observe the animation for 5 seconds ($N = 5000$) resulting in $M = 125$ frames. From these 125 frames we recover frequency content of the periodic signal being observed by enforcing sparsity in reconstruction as described in II-D. We compare *structured sparse* reconstruction on normal camera frames, *normal sparse* and *structured sparse* reconstruction on CSC frames and the results are shown in (d),(e) and (f) respectively. It is important to modulate the scene with a

Condition Number κ (Period P)	$U = 25$	$U = 40$	$U = 47$	$U = 55$	$U = 63$	$U = 91$
NC: largest	1.8×10^{19} (75)	8.6×10^{33} (5)	6.1×10^{32} (47)	4.5×10^{65} (95)	3.4×10^{64} (90)	6.5×10^{48} (70)
CSC: largest	1.3×10^3 (9)	1.4×10^4 (7)	6.0×10^3 (8)	2.1×10^4 (19)	8.1×10^2 (27)	2.4×10^3 (7)
NC: smallest	5.9×10^2 (67)	8.4×10^2 (63)	1.5×10^3 (54)	2.7×10^2 (92)	1.5×10^3 (80)	1.6×10^3 (55)
CSC: smallest	16.5 (67)	11.5 (94)	10.1 (98)	9.7 (90)	10.9 (77)	13.2 (53)

TABLE I: Table comparing the largest and smallest condition numbers of mixing matrix \bar{A} corresponding to normal (NC) and coded strobing exposure (CSC).

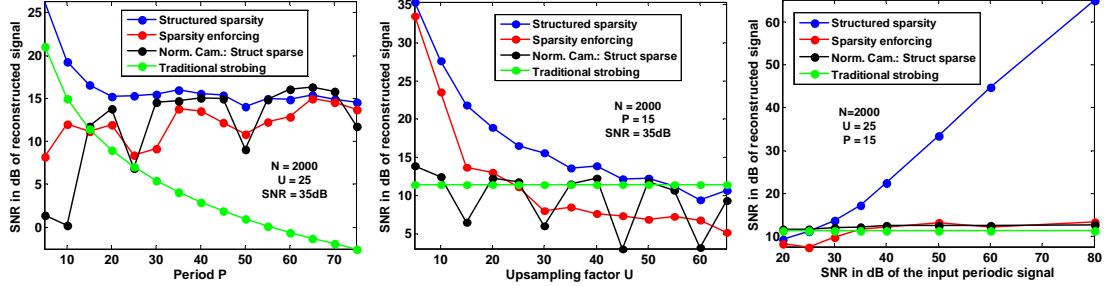


Fig. 9: Performance analysis of structured and normal sparsity enforcing reconstruction for CSC and structured sparsity enforcing reconstruction for normal camera: (a) Reconstruction SNR as the period P increases. (b) Reconstruction SNR as upsampling factor U increases. (c) Reconstruction SNR as the noise in y is varied.

code to capture all frequencies and enforcing both sparsity and structure in reconstruction ensures that the periodic signal is recovered accurately.

Noise analysis and influence of upsampling factor: We perform statistical analysis on the impact of two most common sources of noise in CSC and also analyze the influence of upsampling factor on reconstruction. We recover the signal using *structured sparsity* enforcing reconstruction. First, we study the impact of sensor noise. Figure 11(a) shows the performance of our reconstruction with increasing noise level η . We fixed the upsampling factor at $U = 40$ in these simulations. The reconstruction SNR varies linearly with the SNR of the input signal in accordance with compressive sensing theory. The second most significant source of errors in a CSC are errors in the implementation of the code due to lack of synchronization between the shutter and the camera. These errors are modeled as bit-flips in the code. Figure 11(b) shows the resilience of the coded strobing method to such bit-flip errors. The upsampling factor is again fixed at 40. Finally, we are interested in an understanding of how far the upsampling factor can be pushed without compromising on the reconstruction quality. Figure 11(c) shows the reconstruction SNR as the upsampling factor increases. This indicates that by using structured sparsity enforcing reconstruction algorithm, we can achieve large upsampling factors with a reasonable fidelity of reconstruction. Using the procedure described in previous section we estimate the fundamental frequency as $f_p = 40$ Hz (Figure 11(d)).

IV. EXPERIMENTAL PROTOTYPES

A. Hi-speed video camera

In order to study the feasibility and robustness of the proposed camera, we first tested the approach using a high-speed video camera. We used an expensive 1000 fps video camera, and captured high-speed video. We had to use strong illumination sources to light the scene and capture reasonably noise-free high-speed frames. We then added several of these frames (according to the strobe code) in software to simulate low

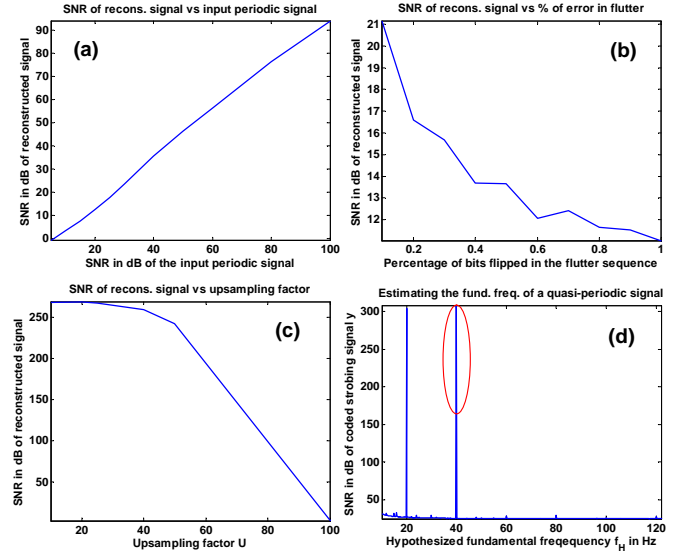


Fig. 11: Performance analysis of CSC: (a) Reconstruction SNR as the observation noise increases. (b) Impact of bit-flips in binary exposure sequence. (c) Coded strobing camera captures the scene accurately upto an upsampling factor $U = 50$. (d) $\|y\|/\|y - \hat{y}\|$ against varying hypothesized fundamental frequency f_H .

speed coded strobing camera frames. The simulated CSC frames were used to reconstruct the high-speed video. Some results of such experiments are reported in Figure 12.

B. Sensor integration mechanism

We implement CSC for our experiments using an off-the-shelf Dragonfly2 camera from PointGrey Research [28], without modifications. The camera allows a triggering mode (Multiple Exposure Pulse Width Mode- Mode 5) in which the sensor integrates the incoming light when the trigger is '1' and is inactive when the trigger is '0'. The trigger allows us exposure control at a temporal resolution of $\delta t = 1$ ms. For every frame we use a unique triggering sequence corresponding to a unique code. The camera outputs the integrated sensor readings as a frame after a specified number of integration

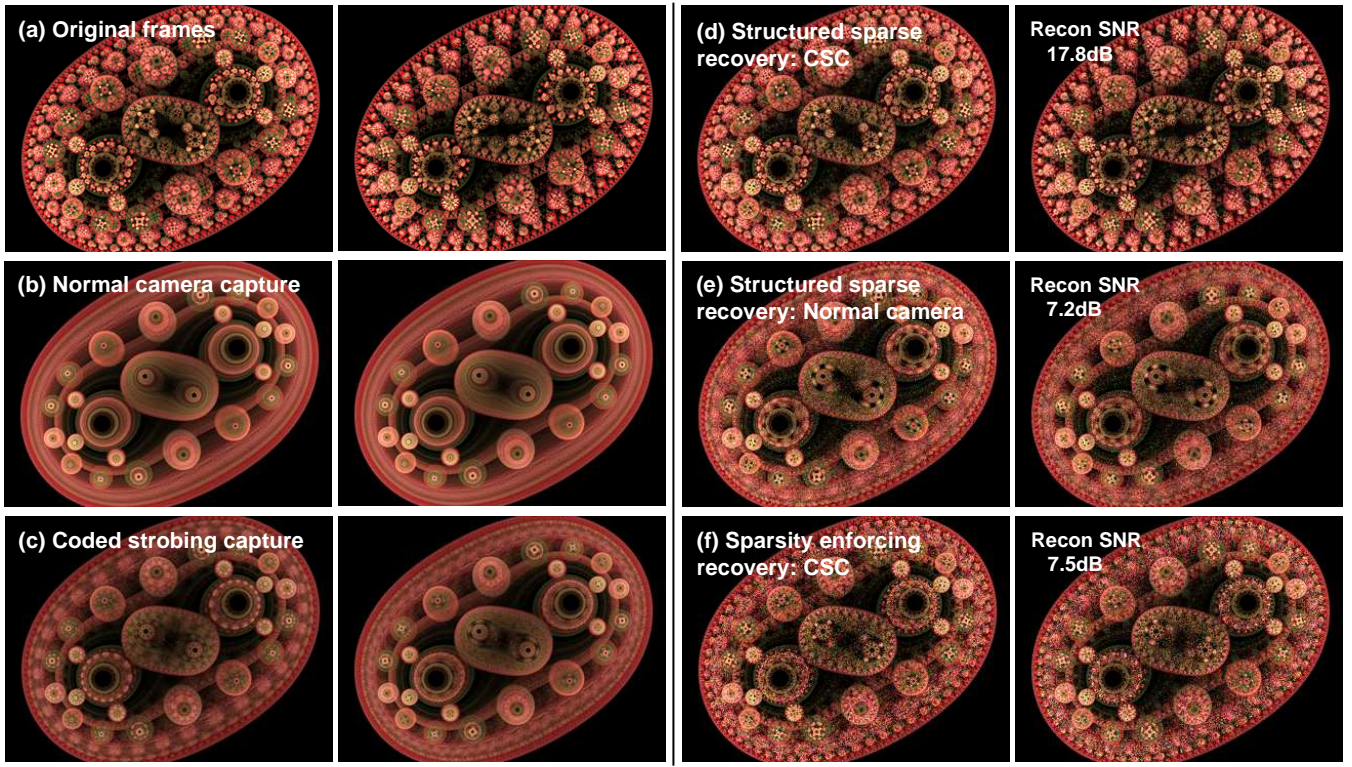


Fig. 10: (a) Original frames of the fractal sequence which repeat every $P = 25$ ms. (b) Frames captured by a normal 25 fps camera. (c) Frames captured by a CSC running at 25 fps. (d) Frames reconstructed by enforcing structured sparsity on CSC frames. (e) Frames reconstructed by enforcing structured sparsity on normal camera frames. (f) Frames reconstructed by enforcing simple sparsity on CSC frames. Overall 5 seconds ($N = 5000$) of the sequence was observed to reconstruct it back fully. Upsampling factor was set at $U = 40$ ($M = 125$) corresponding to $\delta t = 1$ ms. Note that image intensities in (b) and (c) have been rescaled appropriately for better display.

periods. Also, each integration period includes at its end a period of about 30 ms during which the camera processes the integrated sensor readings into a frame. The huge benefit of this setup is that it allows us to use an off-the-shelf camera to slow down high-speed events around us. On the other hand, the hardware bottleneck in the camera restricts us to operate at an effective frame rate of 10 fps (100 ms) and a strobe rate of 1000 strobes/second ($\delta t = 1$ ms).

C. Ferro-electric shutter

The PointGrey Dragonfly2 provides exposure control with a time resolution of 1 ms. Hence, it allows us a temporal resolution of $\delta t = 1$ ms at recovery time. However, when the maximum linear velocity of the object is greater than 1 pixel per ms, the reconstructed frames have motion blur. One can avoid this problem with finer control over the exposure time. For example, a DisplayTech ferro-electric liquid crystal shutter provides an ON/OFF contrast ratio of about 1000 : 1, while simultaneously providing very fast switching time of about $250\mu s$. We built a prototype where the Dragonfly2 captures the frames at usual 25 fps and also triggers a PIC controller after every frame which in turn flutters the ferro-electric shutter with a new code at a specified temporal frequency. In our experiment we set the temporal resolution at $500\mu s$ i.e. 2000 strobes/second.

D. Retrofitting commercial stroboscopes

Another exciting alternative to implement CSC is to retrofit commercial stroboscopes. Commercial stroboscopes used in

laryngoscopy usually allow the strobe light to be triggered via a trigger input. Stroboscopes that allow such an external trigger for the strobe can be easily retrofitted to be used as a CSC. The PIC controller used to trigger the ferro-electric shutter can instead be used to synchronously trigger the strobe light of the stroboscope, thus converting a traditional stroboscope to a coded stroboscope.

V. EXPERIMENTAL RESULTS

To validate our design we conduct two kinds of experiments. In the first experiment, we capture high-speed videos and then generate CSC frames by appropriately adding frames of the high-speed video. In the second set of experiments we captured videos of fast moving objects with a low-frame-rate CSC implemented using a Dragonfly2 video camera. Details about the project and implementation can be found at the webpage <http://www.umiacs.umd.edu/dikpal/Projects/codedstrobing.html>.

A. High-speed video of toothbrush

We capture a high-speed (1000 fps) video of a pulsating Crest toothbrush with quasi-periodic linear and oscillatory motions at about 63 Hz. Figure 4(b) shows the frequency of the toothbrush as a function of time. Notice that even within a short window of 30 seconds, there are significant changes in frequency. We render a 100 fps, 20 fps, 10 fps CSC (i.e., a frame duration of 10 ms, 50 ms, 100 ms respectively) by adding appropriate high-speed video frames, but reconstruct the moving toothbrush images at a resolution of 1 ms as shown

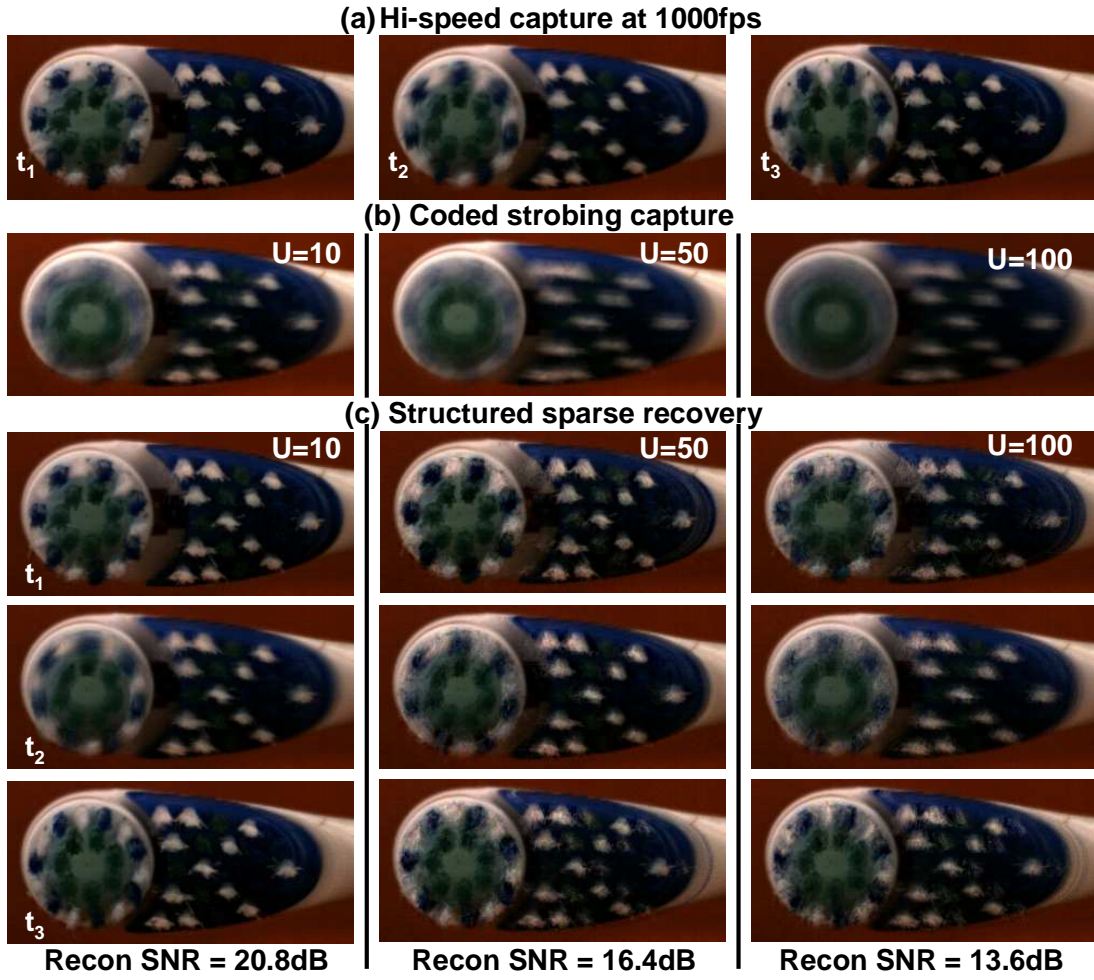


Fig. 12: Reconstruction results of an oscillating toothbrush under three different capture parameters (U): Images for simulation captured by a 1000 fps high-speed camera at time instances t_1, t_2 and t_3 are shown in (a). The second row (b) shows a frame each from the coded strobing capture (simulated from frames in (a)) at upsampling factors $U = 10, 50$, and 100 respectively. Reconstruction at time instances t_1, t_2 and t_3 from the frames captured at $U = 10$ are shown in first column of (c).

in Fig 12c. Frames of the CSC operating at 100, 20 and 10 fps ($U = 10, 50$ and 100 respectively) are shown in Figure 12(b). The fine bristles of the toothbrush add high frequency components because of texture variations. The bristles on the circular head moved almost 6 pixels within 1 ms. Thus the captured images from the high-speed camera themselves exhibited blur of about 6 pixels which can be seen in the recovered images. Notice that contrary to what it seems to the naked eye, the circular head of the toothbrush does not actually complete a rotation. It just exhibits oscillatory motion of 45 degrees and we are able to see it from the high-speed reconstruction.

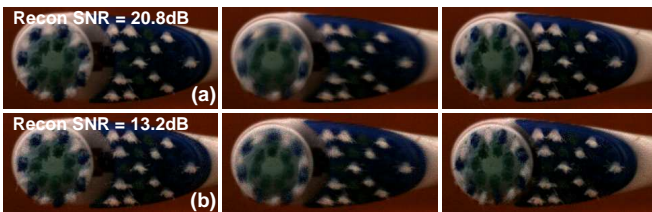


Fig. 13: Reconstruction results of toothbrush with upsampling factor $U = 10$ without and with 15 dB noise in (a) and (b) respectively.

To test the robustness of coded strobing capture and recovery



Fig. 14: Reconstruction results of toothbrush with upsampling factor $U = 50$. Note that using CSC to capture periodic scenes allows us better reconstruction over using a normal camera.

on the visual quality of images, we corrupt the observed images y with white noise having $SNR = 15$ dB. The results of the recovery without and with noise are shown in Figure 13.

We compare frames recovered from CSC to those recovered from a normal camera (by enforcing structured sparsity) to illustrate the effectiveness of modulating the frames. Normal camera doesn't capture the motion in the bristles as well (Figure 14) and is saturated.

B. Mill-tool results using ferro-electric shutter

We use a Dragonfly2 camera with a ferro-electric shutter and capture images of a tool rotating in a mill. Since the tool can

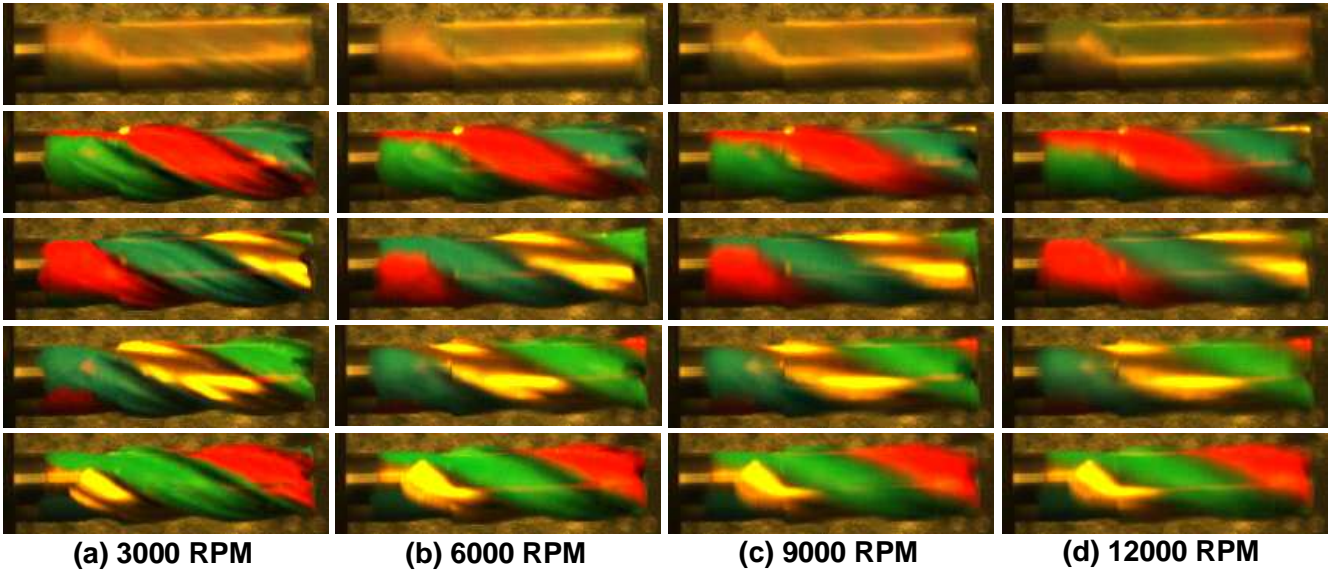


Fig. 15: Tool bit rotating at different rpm captured using coded strobing: Top row shows the coded images acquired by a PGR Dragonfly2 at 25 fps, with an external FLC shutter fluttering at 2000 Hz. (a)-(d) Reconstruction results, at 2000 fps (temporal resolution $\delta t = 500\mu s$), of a tool bit rotating at 3000, 6000, 9000 and 12000 rpm respectively. For better visualization, the tool was painted with color prior to the capture.

rotate at speeds as high as 12000 rpm (200 Hz), to prevent blur in reconstructed images we use the ferro-electric shutter for modulation with a temporal resolution of 0.5 ms. The CSC runs at 25 fps (40 ms frame length) with the ferro-electric shutter fluttering at 2000 strobes/second. Shown in Figure 15 are the reconstructions at 2000 fps ($\delta t = 0.5$ ms) of a tool rotating at 3000, 6000, 9000 and 12000 rpm. Without a priori knowledge of scene frequencies, we use the same strobed coding and the same software decoding procedure for the mill tool rotating at different rpm. This shows that we can capture any sequence of periodic motion with unknown period with a single pre-determined code. In contrast, in traditional strobing prior knowledge of the period is necessary to strobe at the appropriate frequency. Note that the reconstructed image of the tool rotating at 3000 rpm is crisp (Figure 15(a)) and the images blur progressively as the rpm increases. Since the temporal resolution of Dragonfly2 strobe is 0.5 ms, the features on the tool begin to blur at speeds as fast as 12000 rpm (Figure 15(d)). In fact, the linear velocity of the tool across the image plane is about 33 pixels per ms (for 12000 rpm), while the width of the tool is about 45 pixels. Therefore, the recovered tool is blurred to about one-third its width in 0.5 ms.

C. Toothbrush using Dragonfly2 camera

We use a Dragonfly2 camera operating in Trigger Mode 5 to capture a coded sequence of the Crest toothbrush oscillating. The camera operated at 10 fps, but we reconstruct video of the toothbrush at 1000 fps ($U = 100$) as shown in Figure 16. Even though the camera acquires a frame every 100 ms, the reconstruction is at a temporal resolution of 1 ms. If we assume that there are L photons per ms, then each frame of the camera would acquire around $0.5 * 100 * L$ photons. In comparison, each frame of a high-speed camera would accumulate L photons, while traditional strobing camera would accumulate $L * f_P / f_s = 6.3L$ photons per frame.

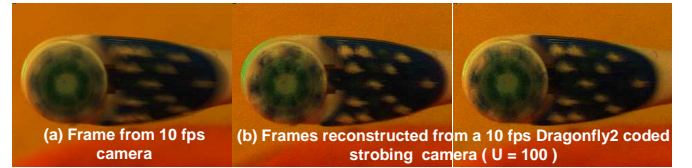


Fig. 16: Demonstration of CSC at upsampling factor $U = 100$ using Dragonfly2. (a) Captured image from a 10 fps CSC (Dragonfly2). (b)-(c) Two reconstructed frames. While the CSC captured an image frame every 100 ms, we obtain reconstructions with a temporal resolution of 1 ms.

D. High-speed video of a jog

Using frames from a high-speed (250 fps) video of a person jogging-in-place we simulate the capture of the scene using a normal camera and the CSC at upsampling factors of $U = 25, 50$ and 75 . The coded frames from CSC are used to reconstruct back the original high-speed frames by enforcing *structured sparsity*. The result of the reconstruction using frames from the CSC is contrasted with frames captured using a normal camera in Figure 17(a). At any given pixel, the signal is highly quasi-periodic since it is not a mechanically driven motion but our algorithm performs reasonably well in capturing the scene. In Figure 17(b) we contrast the reconstruction at a pixel for $U = 25, 50$, and 75 .

VI. BENEFITS AND LIMITATIONS

A. Benefits and advantages

Coded strobing allows three key advantages over traditional strobing: (i) signal to noise ratio (SNR) improvements due to light-efficiency, (ii) no necessity for prior knowledge of dominant frequency, and (iii) the ability to capture scenes with multiple periodic phenomena with different fundamental frequencies.

Light throughput: Light efficiency plays an important role if one cannot increase the brightness of external light sources.

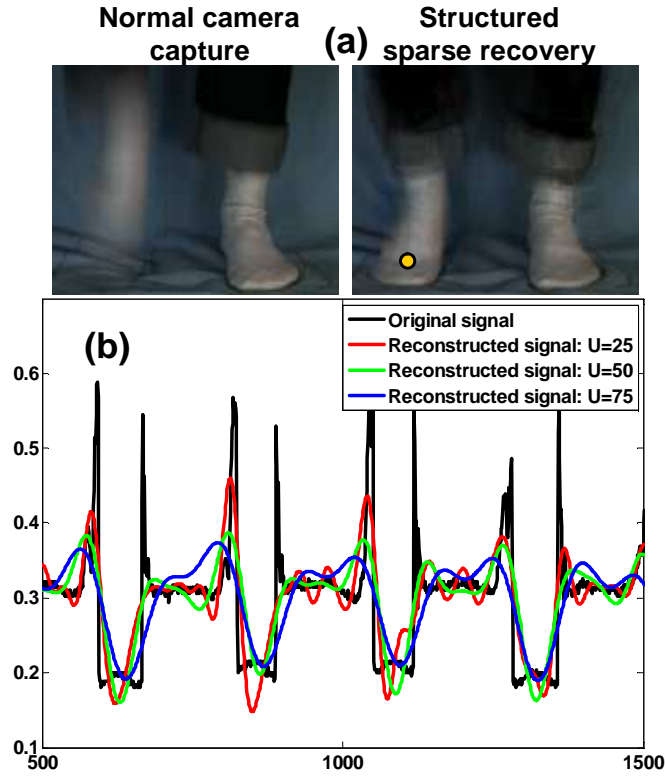


Fig. 17: Frontal scene of a person jogging-in-place. (a) A frame captured by a normal camera (left) and one of the frames recovered from coded strobing capture at $U = 25$ (right). (b) Plot in time of the pixel (yellow) of the original signal and signal reconstructed from coded strobing capture at $U = 25, 50$ and 75 . Note that the low frequency parts of the signal are recovered well compared to the high-frequency spikes.

Let us consider the linear noise model (scene independent) where the SNR of the captured image is given by $LT_{Exposure}/\sigma_{gray}$, where L is the average light intensity at a pixel and σ_{gray} is a signal independent noise level which includes effects of dark current, amplifier noise and A/D converter noise. For both traditional and coded strobing cameras, the duration of the shortest exposure time should at most be $t_{\delta} = 1/(2f_{Max})$. In traditional strobing, this short exposure t_{δ} is repeated once every period of the signal, and therefore the total exposure time in every frame is given by $T_{Strobing} = (1/2f_{Max})(f_P/f_s)$. Since the total exposure time within a frame can be as large as 50% of the total frame duration for CSC, $T_{Coded} = 1/2f_s$. The decoding process in coded strobing introduces additional noise, and this decoding noise factor is $d = \sqrt{\text{trace}((A^T A)^{-1})/M}$. Therefore, the SNR gain of CSC as compared to traditional strobing is given by

$$SNR_{Gain} = \frac{SNR_{Coded}}{SNR_{Strobing}} = \frac{(LT_{Coded})/(d\sigma)}{(LT_{Strobing})/(\sigma)} = \frac{f_{Max}}{df_P} \quad (9)$$

For example, in the case of the tool spinning at 3000 rpm (or 50 Hz), this gain is $20 \log(1000/(2 \cdot 50)) = 20 \text{ dB}$ since $f_{Max} = 1000 \text{ Hz}$ for strobe rate 2000 strobes/second. So coded strobing is a great alternative for light-limited scenarios such as medical inspection in laryngoscopy (where patient tissue burn is a concern) and long range imaging.

Knowledge of fundamental frequency: Unlike traditional strobing, coded strobing can determine signal frequency in post-capture, software only process. This allows for interesting applications such as simultaneous capture of multiple signals with very different fundamental frequencies. Since the processing is independent for each pixel, we can support scenes with several independently periodic signals and capture them without a-priori knowledge of the frequency bands as shown in Figure 18(a). Shown, in Figure 15 are the reconstructions obtained for the tool which was rotating at 3000, 4500, 6000 and 12000 rpm. In all these cases, the same coded shutter sequence was used at capture-time. Also, the reconstruction algorithm can also eminently handle both periodic and quasi-periodic signals using the same framework.

Multiple periodic signals: Unlike traditional strobing, coded strobing allows us to capture and recover scenes with multiple periodic motions with different fundamental frequencies. The capture in coded strobing doesn't rely on frequency of the periodic motion being observed and the recovery of the signal at each pixel is independent of the other. This makes it possible to capture a scene with periodic motions with different fundamental frequency all at the same time using the same hardware settings. The different motions are reconstructed independently by first estimating the respective fundamental frequencies and then reconstructing by enforcing structured sparsity.

We perform experiments on an animation with two periodic motions with different fundamental frequencies. Shown in Figure 18(a) are few frames of the animation with a rotating globe on the left and a horse galloping on the right. The animation was created using frames of a rotating globe which repeats every 24 frames and frames of the classic galloping horse which repeats every 15 frames. For simulation, we assume that a new frame of the animation is being observed at a resolution of $\delta t = 1 \text{ ms}$ and we observe the animation for a total time of 4.8 seconds ($N = 4800$). This makes the period of the globe 24 ms ($f_P = 41.667 \text{ Hz}$) and that of horse 15 ms ($f_P = 66.667 \text{ Hz}$). The scene is captured using a 25 fps ($U = 40$) camera and few of the captured CSC frames are shown in (b). The reconstructed frames obtained by enforcing structured sparsity are shown in (c). Prior to the reconstruction of the scene at each pixel, fundamental frequencies of the different motions were estimated. For one pixel on horse (marked blue in Figure 18(a)) and one pixel on the globe (marked red), the output SNR $\|y\|/\|y - \hat{y}\|$ is shown as a function of hypothesized fundamental frequency f_H in Figure 18(d). The fundamental frequency are accurately estimated as 66.667 Hz for the horse and 41.667 Hz for the globe.

Ease of implementation: The previous benefits assume significance because modern cameras, such as PointGrey Dragon-Fly2, allow coded strobing exposure and hence there is no need for expensive hardware modifications. We transform this off-the-shelf camera instantly into a 2000 fps high-speed camera using our sampling scheme. On the other hand, traditional strobing has been extremely popular and successful because of its direct-view capability. Since our reconstruction algorithm

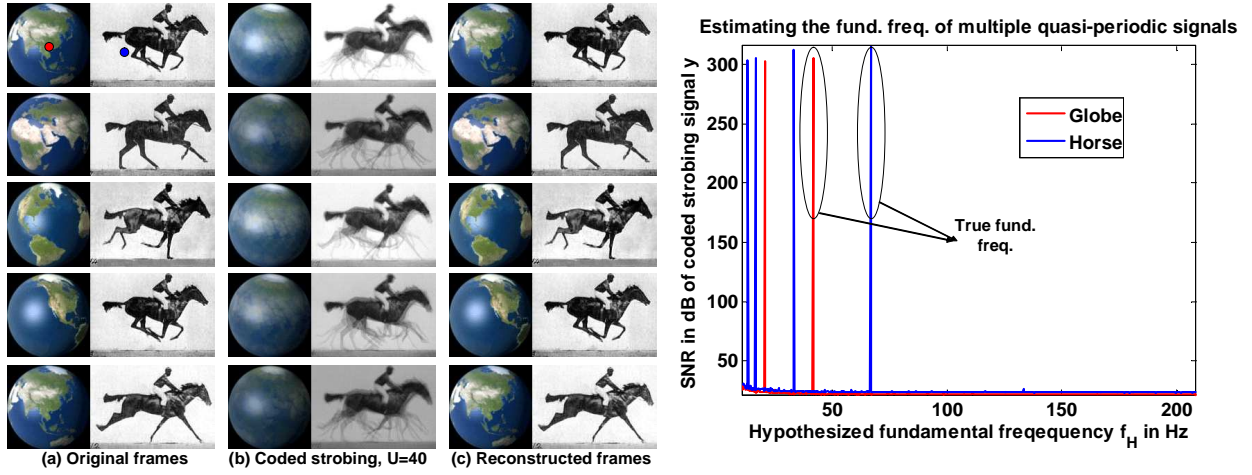


Fig. 18: Recovery of multiple periodic motion in a scene. (a) Periodic events with different periods in the same scene. The scene as captured by CSC with $U = 40$ is shown in (b). The recovered frames are shown in (c). Shown in (d) is the estimated fundamental frequency of globe and horse at points marked red and blue. Note that the last peak in both globe and horse corresponds to the respective fundamental frequency of 41.667 Hz and 66.667 Hz.

is not yet real-time, we can only provide a delayed viewing of the signal. Table II lists the most important characteristics of the various sampling methodologies presented.

B. Artifacts and limitations

We address the three most dominant artifacts in our reconstructions: (a) Blur in the reconstructed images due to time resolution, (b) temporal ringing introduced during deconvolution process, and (c) saturation due to specularities.

Blur: As shown in Fig 19, we observe blur in the reconstructed images when the higher spatio-temporal frequency of the motion is not captured by the shortest exposure time of 0.5 ms. Notice that the blur when $\delta t = 0.5$ ms is less compared to when $\delta t = 1$ ms. The width of the tool is about 45 pixels and the linear velocity of the tool across the image plane is 33 pixels per millisecond. Hence, there is a blur of about 16 pixels in the reconstructed image when $\delta t = 0.5$ ms and 33 pixels when $\delta t = 1$ ms. Note that this blur is not a result of the reconstruction process and is dependent on the smallest temporal resolution. It must also be noted here that while 12000 rpm (corresponding to 200 Hz) is significantly less compared to the 2000 Hz temporal resolution offered by coded strobing, the blur is a result of visual texture on the tool.

Temporal ringing: Temporal ringing is introduced in the reconstructed images during the reconstruction (deconvolution) process. For simplicity, we presented results without any regularization in the reconstruction process (Figure 12(c)). Note that in our algorithm reconstruction is per pixel and the ringing is over time. Figure 20(a) shows temporal ringing at two spatially close pixels. Since the waveforms at these two pixels are related (typically phase shifted), the temporal ringing appears as spatial ringing in the reconstructed images (Figure 16(b)). Either data independent Tikhonov regularization or data dependent regularization (like priors) can be used to improve the visual quality of the reconstructed videos.

Saturation: Saturation in the captured signal y results in sharp edges which in turn leads to ringing artifacts in the recon-

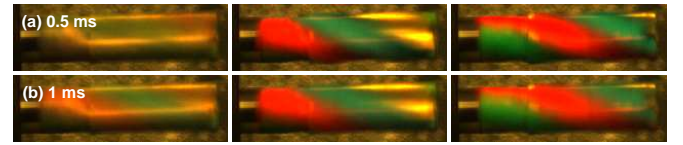


Fig. 19: Coded strobing reconstructions exhibit blur when the temporal resolution δt is not small enough. Shown in (a) and (b) are the same mill tool rotating at 12000 rpm and captured by a strobe with $\delta t = 0.5$ ms and $\delta t = 1$ ms respectively. The reconstructions shown in the second and third column show that $\delta t = 1$ ms strobe rate is insufficient and leads to blur in the reconstructions.

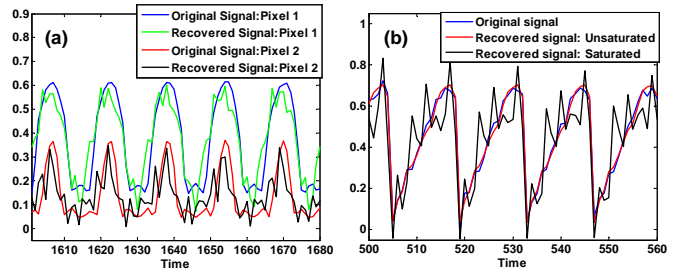


Fig. 20: (a) Ringing artifacts (in time) in the reconstructed signal at two pixels separated by 8 units in Fig 12(c). Also shown are the input signals. Note that artifacts in reconstruction (in time) manifests itself as artifacts in space in the reconstructed image. (b) Artifacts in the reconstructed signal due to saturation in the observed signal y .

structed signal. In Figure 20(b) we can see that the periodic signal recovered from saturated y has temporal ringing. Since reconstruction is independent for each pixel, the effect of saturation is local and does not affect the rest of the pixels in the image. Typical cause of saturation in the captured image is due to specularities in the observed scene. Specularities, that are not saturated, do not pose a problem and are reconstructed as well as other regions.

VII. DISCUSSION AND CONCLUSION

A. Spatial redundancy

In this paper, we discussed a method called coded strobing that exploits the temporal redundancy of periodic signals and

Method	Sampling Rate	Best Scenario	Benefits	Limitations
High-speed (Nyquist)	$2 f_0$	Scene within f_0	Robust	Costly
Strobing (band-pass)	Lower than f_0	Periodic and Brightly lit	Direct-view	Linear search
Non-uniform	Lower than f_0	Brightly lit	No aliasing	Not robust to noise
Coded Strobing	Lower than f_0	Periodic	Light-efficient	No direct-view

TABLE II: Table showing relative benefits and appropriate sampling for presented methods.

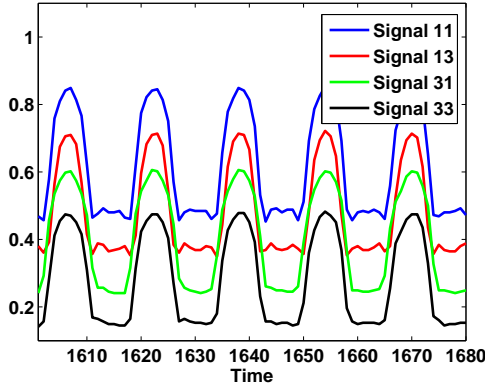


Fig. 21: The waveforms in a neighborhood are highly similar and hence the information is redundant. Shown are the waveforms of 4 pixels at the corners of a 3×3 neighborhood. The waveforms are displaced vertically for better visualization.

in particular, their sparsity in the Fourier domain in order to capture high-speed periodic and quasi-periodic signals. The analysis and the reconstruction algorithms presented considered the data at every pixel as independent. In reality, adjacent pixels have temporal profiles that are very similar. In particular (see Figure 21), the temporal profiles of adjacent pixels are related to each other via a phase shift which depends upon the local speed and direction of motion of scene features. This redundancy is currently not being exploited in our current framework. We are currently exploring extensions of the CSC, that explicitly model this relationship and use these constraints during the recovery process.

B. Spatio-temporal resolution trade-off

The focus of this paper, was on the class of periodic and quasi-periodic signals. One interesting and exciting avenue for future work is to extend the application of the CSC to a wider class of high-speed videos such as high-speed videos of statistically regular dynamical events (e.g., waterfall, fluid dynamics etc) and finally to arbitrary high-speed events such as bursting balloons etc. One alternative we are pursuing in this regard is considering a scenario which allows for spatio-temporal resolution trade-offs, i.e., use a higher resolution CSC in order to reconstruct lower resolution high-speed videos of arbitrary scenes. The spatio-temporal regularity and redundancy available in such videos needs to be efficiently exploited in order to achieve this end.

C. Conclusions

In this paper, we present a simple, yet powerful sampling scheme and reconstruction algorithm that turns a normal video camera into a high-speed video camera for periodic signals. We show that the current design has many benefits over traditional approaches and show a working prototype that is able to turn an off-the-shelf 25 fps PointGrey Dragonfly2 camera into a 2000 fps high-speed camera.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Chellappa for his encouragement and support, John Barnwell for his help with the hardware and Brandon Taylor for making a superb video for the project. Thanks to Amit Agrawal, Jay Thornton and numerous members of Mitsubishi Electric Research Labs for engaging discussions about the project. The work of Dikpal Reddy was supported by the ONR Grant N00014-09-1-0664.

REFERENCES

- [1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1167–1183, 2002.
- [2] R. Baraniuk. Compressive Sensing. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007.
- [3] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *to appear in IEEE Transactions in Information Theory*, 2010.
- [4] B. Basclé, A. Blake, and A. Zisserman. Motion deblurring and super-resolution from an image sequence. In *ECCV '96: Proceedings of the 4th European Conference on Computer Vision-Volume II*, pages 573–582, London, UK, 1996. Springer-Verlag.
- [5] S. Belongie and J. Wills. Structure from periodic motion. In *Workshop on Spatial Coherence for Visual Motion Analysis (SCVMA)*, Prague, Czech Republic, 2004. Springer Verlag, Springer Verlag.
- [6] M. Ben-Ezra and S. Nayar. Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 689–698, 2004.
- [7] I. Bilinskis and A. Mikelson. *Randomized Signal Processing*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1992.
- [8] E. Candes. Compressive sampling. In *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, volume 3, pages 1433–1452, 2006.
- [9] E. Candes and J. Romberg. Sparsity and incoherence in compressive sampling. *INVERSE PROBLEMS*, 23(3):969, 2007.
- [10] E. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, Dec. 2005.
- [11] Checkline. Industrial stroboscopes. <http://www.checkline.com/stroboscopes/>, 2008.
- [12] D. Donoho. Compressed Sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- [13] H. Edgerton. Rapatronic Photographs. http://simplethinking.com/home/rapatronic_photographs.htm, 1951–1963.
- [14] W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65, Mar/Apr 2002.
- [15] H. Greenspan, S. Peled, G. Oz, and N. Kiryati. Mri inter-slice reconstruction using super-resolution. In *MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1204–1206, London, UK, 2001. Springer-Verlag.
- [16] M. G. L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13081–13086, 2005.

- [17] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 53(3):231–239, 1991.
- [18] J. Jia, Y.-W. Tai, T.-P. Wu, and C.-K. Tang. Video repairing under variable illumination using cyclic motions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(5):832–839, May 2006.
- [19] I. Laptev, S. J. Belongie, P. Perez, and J. Wills. Periodic motion detection and segmentation via approximate sequence alignment. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 816–823, Washington, DC, USA, 2005. IEEE Computer Society.
- [20] H. Larsson, S. Hertegård, and B. Hammarberg. Vocal Fold Vibrations: High-Speed Imaging, Kymography, and Acoustic Analysis: A Preliminary Report. *The Laryngoscope*, 110(12):2117, 2000.
- [21] A. Levin, P. Sand, T. Cho, F. Durand, and W. Freeman. Motion-invariant photography. In *Proc. ACM SIGGRAPH*, volume 8, 2008.
- [22] Z. Lin and H.-Y. Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):83–97, 2004.
- [23] F. Marvasti. *Nonuniform Sampling: Theory and Practice*. Kluwer Academic/Plenum Publishers, 2001.
- [24] P. Mergell, H. Herzel, and I. Titze. Irregular vocal-fold vibration High-speed observation and modeling. *The Journal of the Acoustical Society of America*, 108:2996, 2000.
- [25] E. Muybridge. *Animals in Motion*. Dover Publications, 1957. First ed., Chapman and Hall London 1899.
- [26] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301 – 321, 2009.
- [27] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, 1999.
- [28] Pointgrey Research. "PGR IEEE-1394 Digital Camera Register Reference". 2006.
- [29] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph.*, 25(3):795–804, 2006.
- [30] G. Schade, M. Hess, F. Mller, T. Kirchhoff, M. Ludwigs, R. Hillman, and J. Kobler. Physical and technical elements of short-interval, color-filtered double strobe flash-stroboscopy. *HNO*, 50(12):1079–1083, Dec 2002.
- [31] S. Seitz and C. Dyer. View-Invariant Analysis of Cyclic Motion. *International Journal of Computer Vision*, 25(3):231–251, 1997.
- [32] H. S. Shaw and D. D. Deliyski. Mucosal wave: A normophonic study across visualization techniques. *Journal of Voice*, 22(1):23–33, January 2008.
- [33] E. Shechtman, Y. Caspi, and M. Irani. Increasing space-time resolution in video. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 753–768, London, UK, 2002. Springer-Verlag.
- [34] C. Theobalt, I. Albrecht, J. Haber, M. Magnor, and H.-P. Seidel. Pitching a baseball: tracking high-speed motion with multi-exposure images. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 540–547, New York, NY, USA, 2004. ACM.
- [35] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Transactions in Information Theory*, 56(1):520–544, Jan. 2010.
- [36] B. Wandell, P. Catrysse, J. DiCarlo, D. Yang, and A. E. Gamal. Multiple capture single image architecture with a cmos sensor. In *the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, pages 11–17. Society of Multispectral Imaging of Japan, 1999.
- [37] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 765–776, New York, NY, USA, 2005. ACM.



sertation award from the Department of Electrical and Computer Engineering at the University of Maryland. He is a member of the IEEE.



Ashok Veeraraghavan received the Bachelor of Technology degree in electrical engineering from the Indian Institute of Technology, Madras in 2002 and MS and PhD degrees from the Department of Electrical and Computer Engineering at the University of Maryland, College Park in 2004 and 2008, respectively. He is currently a Research Scientist at Mitsubishi Electric Research Labs in Cambridge, Massachusetts. His research interests are broadly in the areas of computational imaging, computer vision and robotics. His thesis received the Doctoral Dissertation award from the Department of Electrical and Computer Engineering at the University of Maryland. He is a member of the IEEE.

Dikpal Reddy received his B.Tech in Electrical Engineering from the Indian Institute of Technology, Kanpur in 2005. He is currently a PhD candidate in the Department of Electrical and Computer Engineering at the University of Maryland at College Park. His research interests are in signal, image and video processing, computer vision and pattern recognition.



Ramesh Raskar joined the Media Lab from Mitsubishi Electric Research Laboratories in 2008 as head of the Labs Camera Culture research group. The group focuses on developing tools to help us capture and share the visual experience. This research involves developing novel cameras with unusual optical elements, programmable illumination, digital wavelength control, and femtosecond analysis of light transport, as well as tools to decompose pixels into perceptually meaningful components. Raskar's research also involves creating a universal platform for the sharing and consumption of visual media.

Raskar received his PhD from the University of North Carolina at Chapel Hill, where he introduced "Shader Lamps," a novel method for seamlessly merging synthetic elements into the real world using projector-camera based spatial augmented reality. In 2004, Raskar received the TR100 Award from Technology Review, which recognizes top young innovators under the age of 35, and in 2003, the Global Indus Technovator Award, instituted at MIT to recognize the top 20 Indian technology innovators worldwide. In 2009, he was awarded a Sloan Research Fellowship. He holds 30 US patents and has received three Mitsubishi Electric Invention Awards. He is currently co-authoring a book on computational photography.