

Per-Pixel Mirror-Based Method for High-Speed Video Acquisition

J. A. S. Lima^a, C. J. Miosso^b, M. C. Q. Farias^c

^a *Department of Computer Science, University of Brasilia, Brasilia, Brazil*

^b *University of Brasilia at Gama, Gama, Brazil*

^c *Department of Electrical Engineering, University of Brasilia, Brasilia – DF, Brazil*

Abstract

High-speed imaging requires high-bandwidth, fast image sensors that are generally only available in high-end specialized cameras. Nevertheless, with the use of compressive sensing theory and computational photography techniques, new methods emerged that use spatial light modulators to reconstruct high-speed videos with low speed sensors. Although these methods represent a big step in the field, they still present some limitations, such as low light efficiency and the generation of measurements with time dependency. To tackle these problems, we propose a per-pixel mirror-based acquisition method that is based on a new kind of light modulator. The proposed method uses moving mirrors to scramble the light coming from different positions, thus ensuring better light efficiency and generating time independent measurements. Our results show that the proposed method and its variations perform better than methods available in the literature, generating videos that are less noisy and that display better content separation.

Keywords: compressive sensing, computational photography, high-speed imaging

1. Introduction

Several applications in science and industry require the acquisition of videos with time resolutions that are from tens to a few hundred times higher than those of typical consumer cameras. Examples include the study of blood flow in cellular structures, imaging of combustion processes, evaluation of precise movements in biomechanical structures, analysis of the mechanics of

novel fluids, detection of movements causing structural fatigue, visual microphones, etc. [1, 2, 3]. In these applications, the most common solution for acquiring high-speed videos involves using special high-speed video cameras, which have highly sensitive sensors capable of acquiring thousands of frames per second (FPS). Unfortunately, the cost of such sensors still prevents their use in most applications. An alternative approach uses a synchronized array of cameras [4, 5], which is also an expensive solution because it requires, typically, 64 to 128 cameras.

A more recent solution consists of using compressive sensing to reconstruct high-speed videos from measurements obtained using sub-60 FPS cameras [6, 7, 8, 9, 10, 11, 12]. Most of these solutions use shutters to scramble the light rays that reach the sensors. One of the devices that uses this approach is the *flutter shutter*, which divides the camera frame time into short-term periods, during which sensors can either receive light or not. Some commercial cameras have implementations of the flutter shutter device [12] that can be used as a compressive sensing acquisition method for reconstructing high-speed periodic scenes [10] and videos with no motion restrictions [11].

Another device that implements this approach is the *per-pixel shutter*, which selects short periods for light exposure using an independent control for each camera sensor. Some compressive sensing high-speed video reconstruction methods [13, 14, 6] are based on per-pixel shutter devices. Although methods that use per-pixel shutters provide better results than methods that use flutter shutters, per-pixel shutters are not currently implemented in commercial cameras. Experimental implementations of per-pixel shutters are performed by attaching an additional optical system to the camera.

Both flutter shutter and per-pixel shutter methods have common drawbacks. First, these methods discard around 50% of the light, reducing light efficiency and consequently image quality after reconstruction. Also, for both methods, the light captured at different time instants is integrated into a single pixel, what means that measurements are time dependent. Therefore, at the reconstruction stage it is difficult to separate the information coming from different time instants. If the goal of an application is to reconstruct high-speed videos, these drawbacks have to be addressed.

Considering the high cost of high-speed cameras and the limitations and drawbacks of acquisition methods that use low-cost cameras, in this paper we propose a new acquisition method for compressive sensing reconstruction of high-speed videos. The proposed method, called *per-pixel mirror-based* (PPM) acquisition method, was described in its initial form in an earlier

publication [15], which contains preliminary results. PPM is based on an acquisition strategy that uses a set of moving mirrors to redirect the light to certain pixels. The method does not discard any light and it separates the temporal information, generating time-independent measurements. In this paper, we detail several variations of the proposed method and compare them with currently available acquisition methods. We tested the proposed method for still images, synthetic videos and natural videos.

The remaining parts of the paper are organized as follows. Sections 2 and 3 briefly describe the compressive sensing theory and the currently available video compressive sensing acquisition methods. Section 4 describes the proposed method, whereas Section 5 presents the simulation results. Finally, Section 6 presents our conclusions and future works.

2. Compressive Sensing

Let N be the dimension of the signal \mathbf{x} to be acquired. This signal \mathbf{x} is said to be sparse in the Ψ domain if only a few projections of \mathbf{x} into the bases of Ψ are non-zero. If the total number of non-zero projections is $K < N$, then the signal is said to be K -sparse. The theory of compressive sensing [16, 17, 18, 19, 20] allows one to acquire and to reconstruct a sparse signal with a smaller number of measurements than the number of samples required by the Nyquist rate. More specifically, compressive sensing allows one to take only M linear measurements from \mathbf{x} , where $M \ll N$, and nonetheless all N components of \mathbf{x} theoretically without any error.

Suppose that \mathbf{x} is K -sparse in the basis Ψ , with $K \ll N$. Let \mathbf{y} be the vector of linear projections of \mathbf{x} into M vectors Φ_i ($i = [1, \dots, M]$). If Φ is the $M \times N$ matrix in which each row is one of those distinct M vectors, then

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s}, \quad (1)$$

where $\Theta = \Phi \Psi$ is an $M \times N$ matrix and $\mathbf{s} = \Psi^{-1} \mathbf{x}$ is the sparse representation of \mathbf{x} in the domain defined by Ψ .

Note that (1) represents the relationship between the available measurements, \mathbf{y} , and the desired signal, \mathbf{x} , which is at first unknown. The acquisition process must provide the components of \mathbf{y} , whereas a reconstruction procedure must provide \mathbf{x} based on \mathbf{y} and on some signal properties. In Section 4, we detail how we obtain \mathbf{y} in our formulation, and we provide the mathematical modelling that relates \mathbf{y} to \mathbf{x} in our proposed methods. In other words,

we describe our particular measurement matrix Φ . For now, we explain the principles based on which we compute \mathbf{x} from \mathbf{y} .

Regarding the reconstruction stage, since (1) is an underdetermined system, there are generally infinite signals \mathbf{s}' that satisfy $\Theta\mathbf{s}' = \mathbf{y}$. Amongst all solutions, we search for the sparsest one. Two properties must be satisfied so that this procedure is stable: the restricted isometry property (RIP) [21] and the incoherence property [22]. According to the RIP, Θ should roughly preserve the lengths of the K -sparse vectors, within a predefined tolerance [22]. Incoherence, on the other hand, requires that the rows of Φ do not have a sparse representation in the Ψ domain [20].

Once the RIP and the incoherence are satisfied, reconstructing a K -sparse signal using these M measurements corresponds to finding the sparsest signal that satisfies these measurements [20]. Directly searching for the sparsest solution, however, is generally unfeasible, as it corresponds to a combinatorial optimization problem [21]. Practical solutions use optimization algorithms based on ℓ_1 - or ℓ_p -minimization problems. Such alternatives reduce the computational complexity, at the cost of increasing the number of linear measurements [16].

In image reconstruction, the ideal problem of finding the sparsest solution by ℓ_0 -minimization can be replaced by the Total Variation (TV) minimization [23, 24], which is the chosen method in this paper. The TV of an $N_1 \times N_2$ image \mathbf{s}' , denoted as $\|\mathbf{s}'\|_{TV}$, is related to the horizontal ($\mathbf{G}_v\mathbf{s}'$) and vertical ($\mathbf{G}_h\mathbf{s}'$) gradients of \mathbf{s}' . If we use ℓ_1 (approach that was already applied in compressive video sensing [25]) to combine $\mathbf{G}_h\mathbf{s}'$ and $\mathbf{G}_v\mathbf{s}'$, TV can be defined as:

$$\|\mathbf{s}'\|_{TV} = \sum_{i=2}^{N_1} \sum_{j=2}^{N_2} |\mathbf{G}_v\mathbf{s}'(i, j)| + |\mathbf{G}_h\mathbf{s}'(i, j)|. \quad (2)$$

The use of TV minimization for image reconstruction is based on the idea that the discrete gradient of natural images tends to generate sparser images. TV can be then viewed as the ℓ_1 of the image in a sparse domain [26]. For image reconstruction, the optimization problem can be described as

$$\hat{\mathbf{s}} = \operatorname{argmin}_{\mathbf{s}'}(\|\mathbf{s}'\|_{TV}), \text{ such that } \mathbf{y} = \Theta\mathbf{s}' = \Phi\Psi\mathbf{s}'. \quad (3)$$

In this equation, Φ is the acquisition matrix and Ψ is the transform basis.

Note that, in our proposed methods, once we obtain the measurements \mathbf{y} described in (1), we apply a numerical optimization procedure such as (2) to compute the desired image \mathbf{x} . Our main contribution is a novel formulation

for obtaining the measurements \mathbf{y} , as described in Section 4, which improves the objective quality in high-speed video reconstruction.

In our approach, we take the measurements in the spatio-temporal domain, i.e., the pixel domain. In other words, Ψ is the identity matrix. Taking $\|\mathbf{s}\|_{TV} = \|\mathbf{D}_i \mathbf{s}\|_1$ and making $\mathbf{s}'' = \mathbf{D}_i \mathbf{s}'$, the optimization problem is then given by

$$\hat{\mathbf{s}} = \operatorname{argmin}_{\mathbf{s}''} (\|\mathbf{s}''\|_1) \text{ such that } \mathbf{y} = \Phi \mathbf{D}_i^{-1} \mathbf{s}''. \quad (4)$$

This method is equivalent to the ℓ_1 minimization problem, where the sparsifying transform is the finite differences operator.

We use TV reconstruction techniques for all tested acquisition methods tested (see Section 5). We refer to the TV reconstruction in 2 spatial dimensions (finite differences of lines and rows in an image) as TV2D and to the TV reconstruction in 2 spatial and 1 temporal dimensions (finite differences among lines, rows, and subsequent subframes) as TV3D. TV2D takes advantage of spatial redundancies, while TV3D takes advantage of spatial and temporal redundancies.

3. Current Video Acquisition Methods

Traditional video acquisition produces frame pictures containing the light captured by sensors during the exposition time. If a long exposure is used, frame pictures may appear blurry for video scenes with a lot of movement (motion blur). However, if we choose a very short exposure time, the picture may appear too dark or noisy, since the amount of light that reaches each sensor is reduced because of the shorter time interval. In order to use compressive sensing reconstruction for obtaining videos with a higher spatial or temporal resolution, we have to linearly combine the scene samples. Without linearly combining samples in a proper way, we cannot guarantee that the incoherence property would be satisfied [27].

In many applications, a higher control of the light flow is desirable. For example, in applications like deblurring [28] or compressive sensing video acquisition [11, 29, 10], it is necessary to start and finish exposure several times during one frame interval. A device that implements this kind of exposure is the flutter shutter (FS), currently available in some commercial cameras. In FS devices, we divide the exposure time into several *equal-time* intervals, which are known as *subframes*. Each subframe can have its exposition activated or not. The incoming light rays corresponding to each activated

subframe are integrated at the end of the frame time. Therefore, linear measurements are obtained from the light acquired from several subframes (at different times).

Figure 1(a) shows the subframes acquired by a high-speed camera, as compared to those acquired by a standard camera in Figure 1(b) and to the FS acquisition process in Figure 1(c). Note that, unless all subframes are activated, some light information is lost. FS is used by Holloway *et al.* in their Flutter Shutter Video Camera [11], achieving a 66% luminous efficiency.

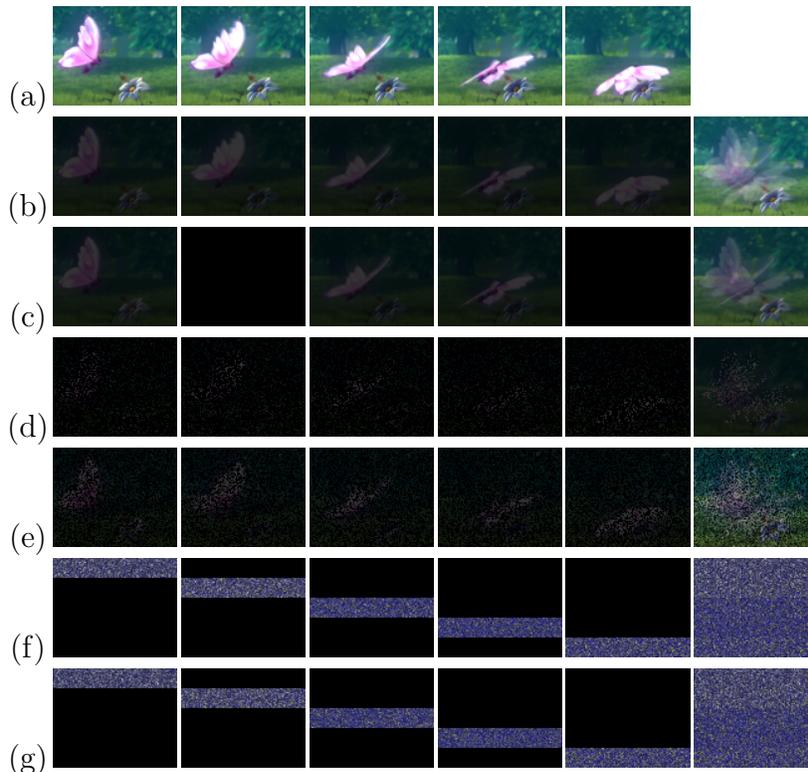


Figure 1: Illustration of light intensity in each subframe for all methods: (a) high-speed camera with sensitive and fast sensors; (b) regular camera; (c) FS; (d) PPS1; (e) PPS2; (f) PPM1; and (g) PPM2. The last column corresponds to the measurements acquired at the end of the frame time, which are later used in reconstruction.

In some applications [30, 13, 14], it is necessary to have an even higher control over the exposure. In these applications, we use a spatial light modulator device (SLD), which provides an independent control of the light flow arriving at each pixel. Currently, one of the devices that implements SLDs

is the digital micromirror device (DMD)[31, 32, 33]. DMDs are implemented using a set of micromirrors that can be set in two positions, allowing two possible direction angles. The first angle makes the lens points the light in the direction of the sensor array, while the second angle makes the lens points the light outwards (i.e. the light is thrown away). Therefore, the light reflection of each mirror can be activated or not. Figure 2 illustrates of the DMD operation. Single pixel camera (SPC) [34] techniques use DMDs with all ‘activated’ mirrors pointing to a single pixel. Figure 2(b) illustrate this process.

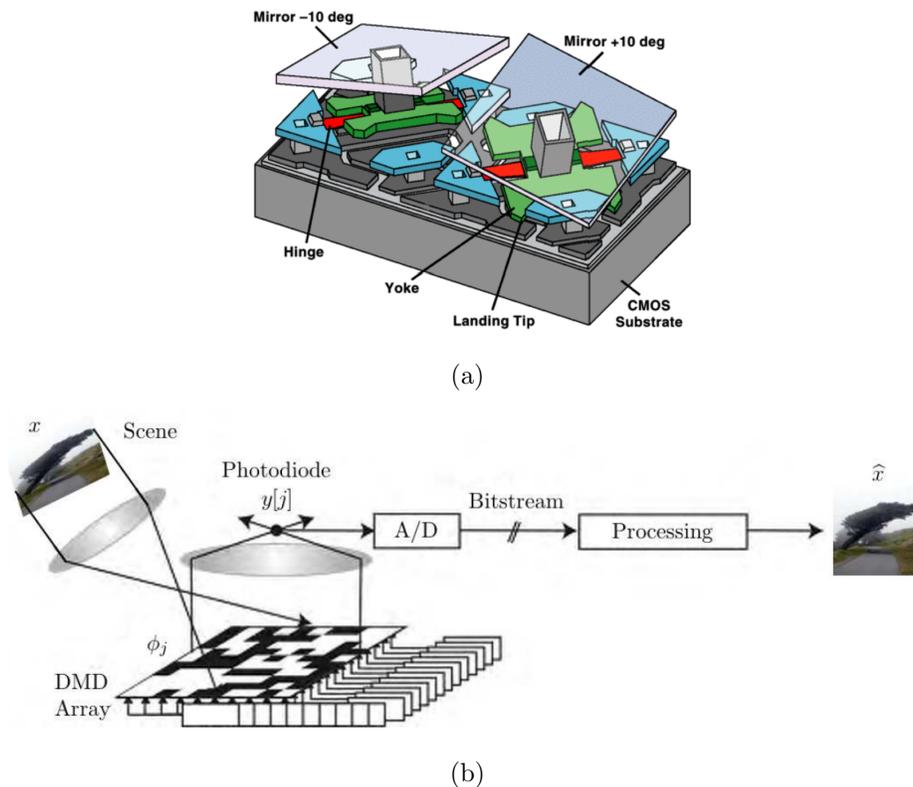


Figure 2: (a) DMD and its 2 mirror direction angles [35]. (b) Use of DMD in SPC method. Adapted from [34].

With SLDs we can implement per-pixel shutter (PPS) devices. PPS is a device that provides a better exposure control than FS devices, since it allows activating or disactivating any of the pixels, several times during the expo-

sure time. Nevertheless, SLDs are still not available in current commercial cameras and experimental implementations use an external optical system. PPS can be used in two ways. In the first one (PPS1), each pixel is activated only once in a subframe. Figure 1(d) illustrates the PPS1 method, showing the linear measurements acquired by the camera sensors. PPS1 is used by Gupta *et al.* in their flexible voxels for motion aware photography [30] and by Hitomi *et al.* for recovering high-speed videos with a training dictionary [14]. One problem of this method is that the majority of the light is not acquired.

In the second form of PPS (PPS2), the pixels can be activated or deactivated regardless of the subframe. At the end of the integration time, the light coming from the pixels activated in different subframes are summed up to produce the acquired pixel. In other words, each acquired pixel contains temporal information from multiple concurrent subframes. Figure 1(e) shows the PPS2 methodology. Reddy *et al.* use PPS2 in its Programmable pixel compressive camera (P2C2) technique [13] for high-speed video recovery. PPS2 implementation produces a luminous loss of around 50 %.

As mentioned earlier, FS and PPS methods have two issues. First, a significant portion of light is not acquired. Second, except for PPS1, at the end of the exposure time acquired pixels contain visual information from several subframes (time dependence). Solving those issues is the motivation for our proposed methods, described next.

4. Per-Pixel Mirror-Based Acquisition Method

In this paper, we propose a new video acquisition method, called *per-pixel mirror-based* (PPM) acquisition method, which generates time-independent measurements and does not discard any light. To implement the method, we propose a new device called high-precision digital micromirror device (HPDMD), which is a higher-accuracy version of the DMD. In the HPDMD, each micro mirror has several possible direction angles, thus allowing the device to point the incoming light to any desired pixel. In addition, similarly to DMD, HPDMD can change the direction angle faster than the FPS. HPDMD is able to simulate the behavior of a regular DMD and, therefore, can be used to simulate all acquisition methods presented in this paper.

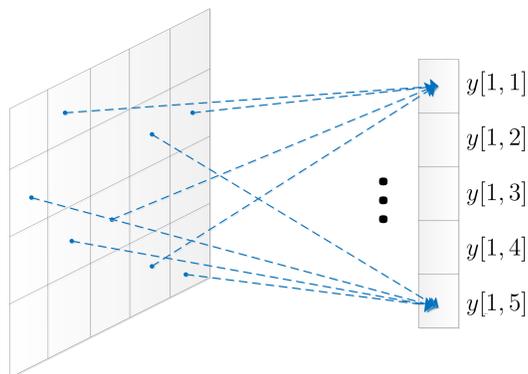
The proposed method operates as follows. For every subframe, light rays coming from different pixels are redirected to common sensors, and each pixel contributes to exactly one CS measurement. In this manner, let k be number of subframes, while N_1 and N_2 are the number of camera pixels in

the horizontal and in the vertical directions, respectively. Also, let N be the total dimension of the signals to be reconstructed, meaning the number of camera pixels multiplied by the number of subframes, so that $N = kN_1N_2$. For each subframe, $(N_1N_2)/k$ measurements are taken, a choice that results in the total number of measurements for all subframes (M) matching the number of camera pixels.

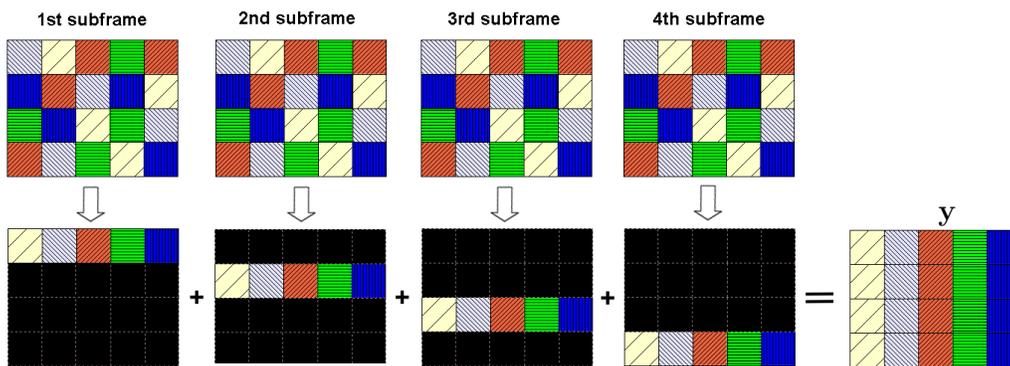
In this scheme, note that since each subframe results in $(N_1N_2)/k$ measurements, and since each of the N_1N_2 pixels contributes to exactly one measurement, each measurement corresponds to the sum of the contents of k pixels. Hence, there are k mirrors to redirect the light that composes a single measurement. There is a choice, however, of which pixels contribute to a same measurement, and a choice of whether the configuration of mirrors changes from one subframe to the next or not. In either case, for each subframe the combinations of pixels that compose the $(N_1N_2)/k$ measurements is based on a random pool, each measurement combining a random combination of pixels.

Figure 3 illustrates the case in which the mirror configuration is fixed amongst subframes, hereby denominated PPM1. In this example, there are $k = 4$ subframes, $N_1 = 4$ rows and $N_2 = 5$ columns. As shown in Figure 3(a), for each subframe, light from $k = 4$ pixels is redirected by k mirrors, in order to compose each of the $(N_1N_2)/k = 5$ measurements for that subframe. Figure 3(b) summarizes the measurement procedure, with similar image patterns representing pixels that contribute to the same measurement. Note that the total number of measurements matches the number of camera pixels (N_1N_2), even though $N > N_1N_2$. However, no light is discarded since all light arriving at each camera pixel is redirected in order to compose a measurement. Furthermore, all light originating from a single subframe is directed into a single set of N/k pixels. This means that light from a specific subframe is not mixed with light coming from another subframe, i.e., there is no time dependence between measurements. On the other hand, there is spatial dependence among samples.

Note that the fact that all measurements in Figure 3 form a single row in \mathbf{y} , for each frame, is just a consequence of the number of subframes (k) matching the number of camera rows (N_1), in this simple example. In this case, the number of measurements in a subframe, given by $(N_1N_2)/k = (N_1N_2/N_1)$, matches the number of camera columns (N_2), which means that a single row in \mathbf{y} is generated per subframe. In Figure 4, we show an example with $N_1 = 4$, $N_2 = 5$ and $k = 2$. In this case, there are $(N_1N_2)/k = 10$ measurements for



(a)



(b)

Figure 3: Proposed method in the PPM1 configuration, with $N = 80$ (20 pixels times 4 subframes), $M = 20$ (number of measurements), and $k = 4$ (number of subframes). (a) Example of the measurement process for each subframe, with randomly selected pixels combining into common measurements through a HPDMD device; the mirror configuration remains the same for all subframes. (b) Summary of the resulting measurements, which are combinations of sets of pixel contents taken from each subframe independently. Each image pattern represents a set of pixels that combine into a single measurement, for each subframe.

each subframe, so that each subframe generates two rows in \mathbf{y} .

Based on this description, we can infer the equations that describe the taken measurements in PPM1, as a function of the pixels contents in each subframe. Let \mathbf{P}^i represent the 2-column matrix in which the rows contain the coordinates of all the pixels that compose the i^{th} measurement,

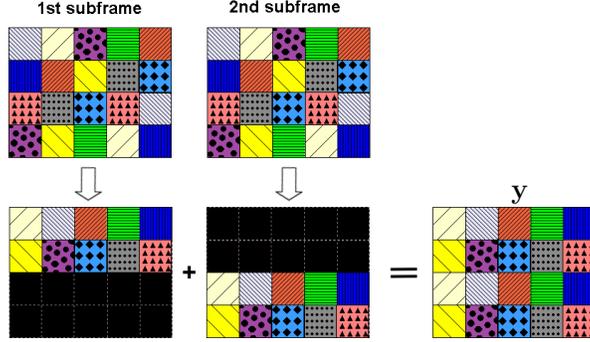


Figure 4: Example of the measurement process for the proposed method in the PPM1 configuration, with $N = 40$ (20 pixels times 2 subframes), $M = 20$ (number of measurements), and $k = 2$ (number of subframes). Each image pattern represents a set of pixels that combine into a single measurement, for each subframe. Note that in this case there are 10 measurements per subframe, so that each subframe generates two rows in \mathbf{y} .

$\forall i \in \{1, 2, \dots, (N_1 N_2)/k\}$. In the case of Figure 3,

$$\mathbf{P}^1 = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}, \quad \mathbf{P}^2 = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 5 \\ 4 & 2 \end{bmatrix}, \quad \mathbf{P}^3 = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 2 & 2 \\ 4 & 1 \end{bmatrix},$$

$$\mathbf{P}^4 = \begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}, \quad \mathbf{P}^5 = \begin{bmatrix} 2 & 1 \\ 2 & 4 \\ 3 & 2 \\ 4 & 5 \end{bmatrix}. \quad (5)$$

Also, let $x[i, j, f]$ represent the pixel content in row i and column j for the f^{th} subframe in the video \mathbf{x} that we want to reconstruct. In this case, the measurement $y[m_1, m_2]$ is given by

$$y[m_1, m_2] = \sum_{n=1}^k x \left[P^{m_2}(n, 1), P^{m_2}(n, 2), Q \left(m_1 - 1, \frac{N_1}{k} \right) + 1 \right], \quad (6)$$

which corresponds to the linear combination exemplified in Figure 3, with $Q(a, b)$ being the highest integer not greater than a/b .

The second form of our proposed method uses a different mirror configuration for every subframe, based on a random pool and redirecting light from

each pixel in order to compose a single measurement for each subframe. Figure 5 illustrates this configuration, which is referred to as PPM2. Differently from PPM1, in PPM2 each subframe has an individual random pattern of k mirrors. Therefore, this configuration is often more interesting, because it better satisfies the incoherence property.

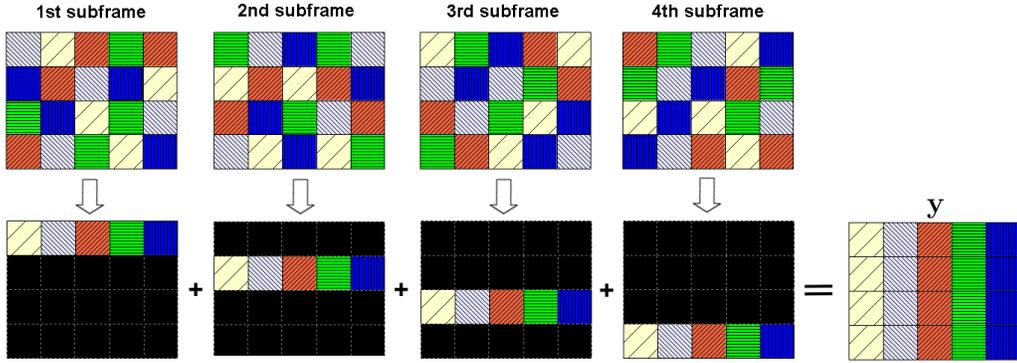


Figure 5: Proposed method in PPM2 configuration with $N = 80$, $M = 20$, and $k = 4$. For each subframe, $M/k = 5$ measurements are taken, corresponding to sums of individual pixel contents. The mirror pattern is different for all subframes, so that the combinations of pixels change from subframe to subframe. Each image pattern represents a set of pixels that combine into a single measurement, for each subframe.

In PPM2, let $\mathbf{P}^{i,f}$ represent the 2-column matrix in which the rows contain the coordinates of all the pixels that compose the i^{th} measurement for subframe f , $\forall i \in \{1, 2, \dots, (N_1 N_2)/k\}$. For the example of Figure 5, we have

$$\begin{aligned}
 \mathbf{P}^{1,1} &= \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}, \mathbf{P}^{2,1} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 5 \\ 4 & 2 \end{bmatrix}, \mathbf{P}^{3,1} = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 2 & 2 \\ 4 & 1 \end{bmatrix}, \mathbf{P}^{4,1} = \begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}, \mathbf{P}^{5,1} = \begin{bmatrix} 2 & 1 \\ 2 & 4 \\ 3 & 2 \\ 4 & 5 \end{bmatrix}, \\
 \mathbf{P}^{1,2} &= \begin{bmatrix} 2 & 1 \\ 2 & 3 \\ 4 & 2 \\ 4 & 4 \end{bmatrix}, \mathbf{P}^{2,2} = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 3 & 4 \\ 4 & 1 \end{bmatrix}, \mathbf{P}^{3,2} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \\ 3 & 1 \\ 3 & 5 \end{bmatrix}, \mathbf{P}^{4,2} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 3 & 3 \\ 4 & 5 \end{bmatrix}, \mathbf{P}^{5,2} = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 3 & 2 \\ 4 & 3 \end{bmatrix}, \\
 \mathbf{P}^{1,3} &= \begin{bmatrix} 1 & 1 \\ 1 & 5 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}, \mathbf{P}^{2,3} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \\ 3 & 2 \\ 4 & 5 \end{bmatrix}, \mathbf{P}^{3,3} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 1 \\ 4 & 2 \end{bmatrix}, \mathbf{P}^{4,3} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 3 \\ 4 & 1 \end{bmatrix}, \mathbf{P}^{5,3} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 5 \\ 4 & 4 \end{bmatrix}, \\
 \mathbf{P}^{1,4} &= \begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}, \mathbf{P}^{2,4} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 5 \\ 4 & 2 \end{bmatrix}, \mathbf{P}^{3,4} = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 4 & 3 \\ 4 & 5 \end{bmatrix}, \mathbf{P}^{4,4} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 2 & 5 \\ 3 & 4 \end{bmatrix}, \mathbf{P}^{5,4} = \begin{bmatrix} 1 & 5 \\ 2 & 3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix}. \quad (7)
 \end{aligned}$$

Then, the measurement $y[m_1, m_2]$ in the case of PPM2 is given by

$$y[m_1, m_2] = \sum_{n=1}^k x [P^{m_2, f}(n, 1), P^{m_2, f}(n, 2), f], \quad (8)$$

where

$$f = Q \left(m_1 - 1, \frac{N_1}{k} \right) + 1. \quad (9)$$

Once we acquire the measurements described by (8), we proceed to the reconstruction stages using the TV minimization methods as in (4).

5. Simulation Tests

In this paper, we compare the performance of three families of acquisition methods: FS, PPS, and PPM. Table 1 shows the main properties of these three families of acquisition methods, more specifically, we present information about the viability, dependency, and amount of light information loss for each method.

Table 1: Properties of tested acquisition methods.

Acquisition Method	Viability in video cameras	Dependency of measurements	light information
Traditional	All cameras	temporal	100%
FS	Some cameras	temporal	50% to 67%
PPS1	Requires DMD	none	$100/k\%$
PPS2	Requires DMD	temporal	$\sim 50\%$
PPM1	Requires HPDMD	spatial	100%
PPM2	Requires HPDMD	spatial	100%

The information in Table 1 is based on the previous description of the proposed methods (in the case of PPM1 and PPM2) as well as on the description of the traditional, FS, PPS1 and PPS2 methods already discussed in Figure 1. In fact, Figure 1 shows several images illustrating the level of light intensity acquired in each subframe for each one of the methods. Each column in the figure corresponds to the measurements in each subframe, while the last column shows measurements obtained for the complete frame time acquisition. Figure 1(a) shows a high-speed camera acquisition, in which

each subframe is acquired in full-resolution. As expected, the resulting images have good light intensity levels and the movement of the video is well represented. Figure 1(b) shows the acquisition using a regular camera. At the end of the frame time (last column), we can notice that there is a motion blur effect, as a result of the integration of the movement in the frame.

Figure 1(c) shows the FS acquisition method. The light intensity is proportional to the number of chosen subframes, which in this case is 60%. Figure 1(d) shows the PPS1 acquisition method. Notice that, in this case, a small amount of light information is acquired (only 20%). Figure 1(e) shows the PPS2 acquisition method. In this case, more light is acquired than for PPS1, but the measurements are scrambled. Figure 1(f) shows the PPM1 acquisition method, while Figure 1(g) shows PPM2. In these two methods, for each subframe, all light is concentrated on pixels of a portion of the video frame area. These portions of pixels do not overlay, guaranteeing time independent measurements.

In the next sections, we evaluate the proposed method using 3 sets of tests: (1) still images, (2) synthesized videos, and (3) natural videos. For each set, we start with an original image or video, simulate the acquisition for each method and reconstruct the signal. Then, we compare the Signal-to-Noise Ratio (SNR) between reconstructed and original signals.

5.1. Part 1: Test on Natural Images

The first set of tests were conducted with still (natural) images. We use a set of gray levels images, with three image sizes: 64×64 ($N = 4,096$), 256×256 ($N = 65,536$) and $1,024 \times 1024$ ($N = 1,048,576$). For an image with a total of N pixels, we consider a Sub-sampling ratio of k , what implies that N/k samples are acquired from each image. This procedure simulates the acquisition of one subframe for a exposure time of $1/k$ of the frame time. In this paper, we show only the results for $k = 4$, which are representative of the performance for other values of k .

We compare the results of acquisition methods, after reconstruction with TV2D. In this set of tests, we consider only acquisition methods that produce temporally independent measurements, i.e. PPS1 and PPM. For comparison purposes, we also use a simple interpolation reconstruction. Figure 6 shows the results for a 64×64 image, while Figure 7 shows zoomed details of the results for a 1024×1024 image. Table 2 shows the SNR values for each test. We can observe that the SNR values corresponding to interpolation results are not much different from what was obtained with PPS1, although

the image reconstructed with PPS1 (Figure 6(c)) has more details than the one reconstructed with interpolation (Figure 6(b)). The reconstruction using PPM has the best performance, both qualitatively and quantitatively.

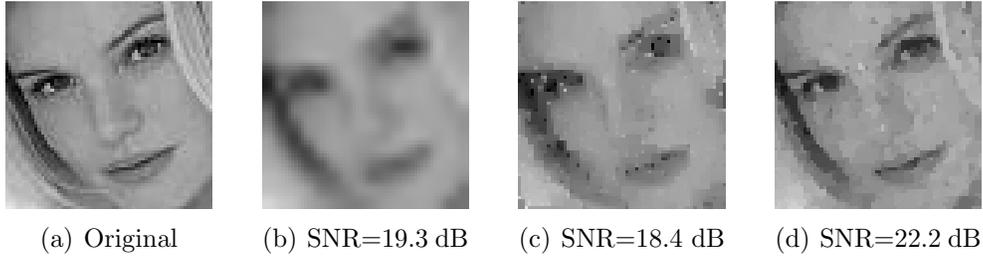


Figure 6: Reconstruction results for a 64×64 image. (a) Original image. (b) Equally spaced sub-sampled image reconstructed with interpolation, SNR=19.3 dB. (c) PPS1 (+TV2D) reconstruction, SNR=18.4 dB. (d) PPM (+TV2D) reconstruction, SNR=22.2 dB.

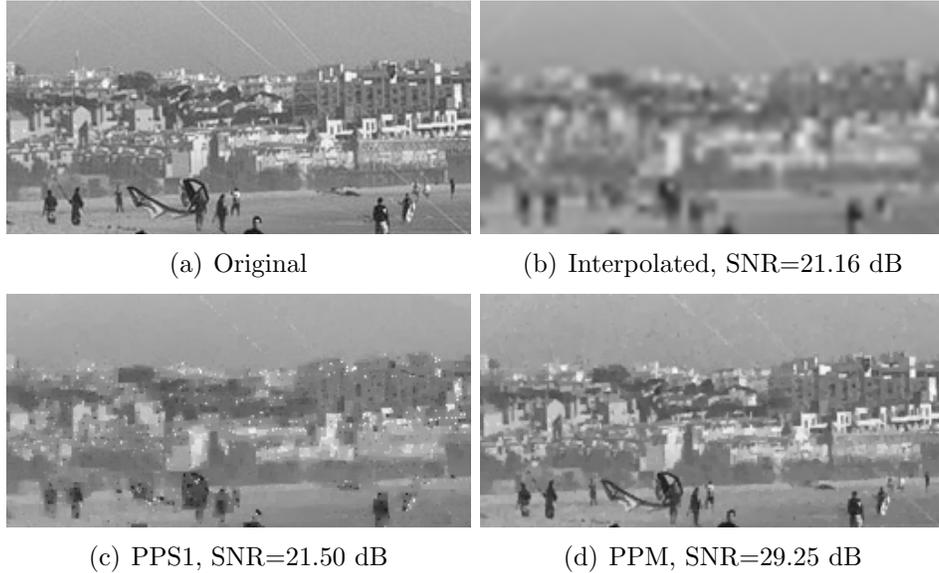


Figure 7: Reconstruction results for an excerpt of a 1024×1024 image. (a) Original image. (b) Equally spaced sub-sampled image, reconstructed by interpolation, SNR=21.4 dB. (c) PPS1 (+TV2D) reconstruction, SNR=21.5 dB. (d) PPM (+TV2D), SNR=29.3 dB.

For larger images, the performances of PPS1 and PPM increase. In fact, the larger the image, the larger the difference in performance for results

Table 2: SNR of images reconstructed using interpolation, PPS1, and PPM ($k = 4$).

Image Size	SNR (dB)		
	Interpolation	PPS1 (+TV2D)	PPM (+TV2D)
64×64	19.3	18.4	22.2
256×256	16.5	16.0	23.2
1024×1024	21.2	21.5	29.3

of different acquisition methods. This can also be observed by comparing the images in Figures 6 and 7. In particular, notice that the image shown in Figure 7(d), which was reconstructed with PPM from only 1/4 of the measurements, has a good image quality showing very small details of the original image.

The results of this initial test show the potential of the proposed method for high temporal resolution video. If we divide the frames of a camera in 4 and take linear measurements from each subframe, we are able to recover four images with a good quality. Therefore, we can increase the temporal resolution 4 times, maintaining the original spatial resolution. There is temporal redundancy among subframes, what means that the four subframes can be completely different from each other and the results will not be affected.

Since videos often have a lot of temporal redundancy, if the temporal redundancy is explored, we can get even better results. In the next section, we present the results of applying the proposed acquisition methods for video signals, taken into account the temporal redundancy between video frames.

5.2. Part 2: Tests with Synthetic Videos

In this section, we test the acquisition methods using a set synthetic videos, which simulate several aspects of objects in a natural video, like movement, occlusion, deformation, and changes in light intensity. We use ellipses and rectangles as scenes objects. These type of videos are based on the Shepp-Logan phantom [36], that is widely used in compressive sensing in medical applications, showing a good performance with the TV reconstruction. Figure 8(a) shows a sample synthetic frame used in our tests. For the five acquisition methods, we use videos of sizes equal to $100 \times 100 \times 128$, $100 \times 100 \times 256$, $200 \times 200 \times 128$, and $200 \times 200 \times 256$. We tested the methods with Sub-sampling ratios ($1/k$) equal to 50%, 25%, 12.5%, and 6.25% s. For

all methods and Sub-sampling ratios, we performed the reconstruction using the TV3D minimization.

Table 3: Mean SNR (dB) of videos reconstructed with FS, PPS1, PPS2, PPM1, and PPM2.

Video size	Acquisition method	Sub-sampling ratio ($1/k$)			
		$k = 2$	$k = 4$	$k = 8$	$k = 16$
100× 128	FS	42.9	6.3	2.4	1.7
	PPS1	9.2	3.0	-0.3	-1.7
	PPS2	9.3	11.1	6.7	5.3
	PPM1	27.2	20.1	12.2	1.2
	PPM2	67.2	55.6	33.7	8.7
100× 256	FS	42.4	8.5	1.2	0.74
	PPS1	9.2	2.7	-0.3	-1.7
	PPS2	9.3	11.6	7.0	5.3
	PPM1	31.2	18.6	11.5	1.9
	PPM2	76.8	55.7	32.9	8.8
200× 128	FS	49.5	9.0	4.9	4.8
	PPS1	11.8	12.0	0.8	-1.4
	PPS2	11.7	13.8	10.4	8.4
	PPM1	33.7	27.4	24.9	8.2
	PPM2	71.3	55.2	44.3	22.6
200× 256	FS	36.8	9.0	6.8	1.9
	PPS1	11.8	12.0	0.8	-1.4
	PPS2	11.7	14.1	10.4	8.5
	PPM1	48.3	33.2	25.9	6.4
	PPM2	73.1	58.8	44.9	25.6

SNR results are presented in Table 3. Note that, in most cases, the reconstructions using PPM (both PPM1 and PPM2) have the best performances, with gains of up to two dozens dB in SNR, when compared to PPS2 (the best PPS variation). When we fix the spatial resolution and vary the number of frames, results do not change much. However, when we increase the spatial resolution and fix the number of frames, SNR values increase considerably. This happens for all acquisition methods and all Sub-sampling ratios, being more evident for the proposed method and for higher Sub-sampling ratios.

It could be argued that the increase in performance for higher spatial resolutions is due to an increase in the number of pixels, i.e. by increasing

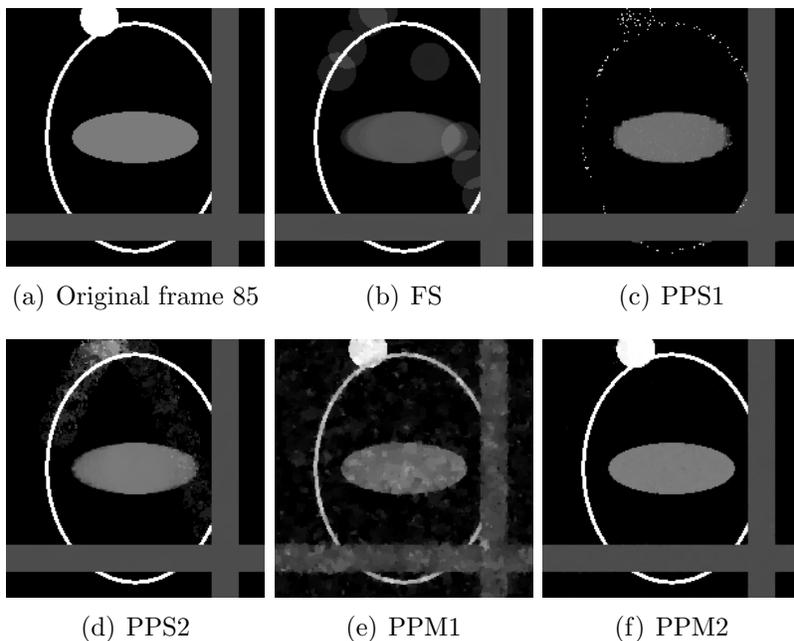


Figure 8: Reconstruction results for methods FS, PPS1, PPS2, PPM1 and PPM2, considering a $200 \times 200 \times 128$ synthetic video and a Sub-sampling ratio of $1/16$.

either the spatial resolution or the number of frames, we are increasing the number of pixels. However, the first set of videos has size $100 \times 100 \times 128 = 1.28$ million pixels, the second has size $100 \times 100 \times 256 = 2.56$ million pixels, the third has $200 \times 200 \times 128 = 5.12$ million pixels and the fourth, $200 \times 200 \times 256 = 10.24$ million pixels. Nevertheless, although the second set of videos has two times more pixels than the first set, the quality differences for the two sets is small. The same happens between the fourth and third set of videos. On the other hand, although the third set has twice the number of pixels as the second set, its SNR value is much larger. So, the spatial resolution of the video seems to be more important for quality than the total number of pixels in the video.

Figure 8 shows the acquisition results for the 5 tested methods, using a Sub-sampling factor equal to $1/16$. The images correspond to the 85-th subframe of a $200 \times 200 \times 128$ synthetic video. Figure 8(a) shows the original subframe, which is the 5-th frame of a set of 16 frames that were reconstructed together (frames 81 to 96). In this set of frames, the white circle (shown in the top of the image) moves to the top region of the image, hits the top edge,

and then moves to the bottom right region. This particular frame (the 85th) shows the moment when the white circle is at the edge.

The quality of the FS frame reconstruction, shown in Figure 8(b), is not very good showing an ‘interference’ of the adjacent frames. PPS1 frame reconstruction (see Figure 8(c)) does not show traces of adjacent (previous or subsequent) frames, since in PPS1 there is no temporal dependencies among measurements corresponding to different subframes. However, a lot of original information is missing, since the amount of acquired light is too low for an accurate scene reconstruction. The PPS2 frame reconstruction, shown in Figure 8(d), shows traces of adjacent subframes. This shows that time dependence in measurements may influence the quality of the reconstruction, especially for areas where there is scene movement. We observe that the higher the Sub-sampling ratio, the greater this dependence and, therefore, the more traces of adjacent frames are present in the reconstructed frame.

PPM shows how these two problems can be addressed simultaneously. Figures 7(e) and 7(f) show frame pictures reconstructed using the PPM1 and PPM2 variations of the proposed method, respectively. We observe that, in these reconstructed frames, there are no traces of the adjacent subframes. In addition, since 100% of light is acquired, the scene has more of the original content than results obtained with PPS1. The quality of the PPM2 reconstruction (Figure 8(f)) is better than the quality of the PPM1 reconstruction (Figure 8(e)), showing that the use of a random mirror pattern for each sub-frame leads to better results. Overall, we see that methods with temporal independence are able to better separate the temporal information.

5.3. Part 3: Tests with Natural Videos

The synthetic (phantom based) videos used in the previous section are very sparse with respect to the finite difference domain, as measured by the TV operator. So, TV minimization results are, in general, good. Unfortunately, natural videos are not as sparse as phantom based videos. In this section, we present results obtained testing the acquisition methods on natural videos. Since previous results show that spatial resolution has a greater impact on the reconstruction quality than the number of frames, we set the number of frames to a minimum value and varied the spatial resolution. Since the number of frames depends on the Sub-sampling ratio, to test a Sub-sampling ratio of 1/16, we need at least 16 video frames.

The 12 original videos used in this test are from The Consumer Digital Video Library (CDVL). The videos are uncompressed (.avi format) with

resolution 720×1280 , with 50 FPS and 4:2:0 color sampling. For these videos, we choose 1 set of 16 frames and converted them to gray levels. Figure 9 shows the first and last frame of each of the 16 frames selected. The greater the difference between the first and last frame, the higher the temporal activity of the video. For the test, it is important that the videos are high quality, containing no visible degradations.

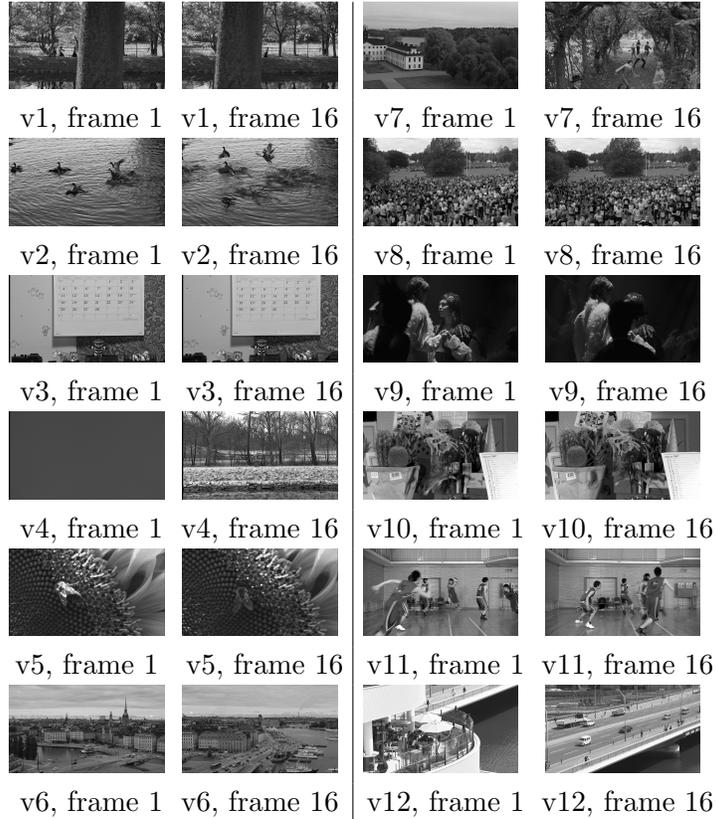


Figure 9: First and last frames of each of the 12 original videos.

In the tests performed on phantoms, the best results for the higher Sub-sampling ratios were obtained for PPS2 and PPM2. Since the tests on HD content require a considerable amount of time and disk space, we choose not to re-test the FS, PPS1 and PPM1 methods. But, as our method allows a frame-by-frame reconstruction, we added two variations to the proposed methodology. First, with the same PPM2 measurements, we perform a TV2D reconstruction of each frame individually. We take the first $M/16$ samples

Table 4: Mean SNR (dB) for natural videos reconstructed with PPS and PPM methods, with a Sub-sampling ratio of 1/16.

Video	PPS2			PPM2			PPM-2D			PPM3		
k	4	8	16	4	8	16	4	8	16	4	8	16
1	11.3	10.0	8.4	12.7	9.1	7.3	15.0	12.4	10.7	15.8	13.2	11.5
2	15.1	13.4	11.9	17.4	12.3	10.0	20.2	17.0	14.7	20.4	17.5	15.3
3	18.4	17.9	14.6	21.7	13.9	11.1	27.7	23.9	21.3	27.1	24.0	21.7
4	12.1	10.2	9.2	15.0	12.1	10.4	17.0	14.7	13.3	17.2	14.9	13.5
5	13.3	12.0	10.1	18.9	11.8	8.4	30.2	25.9	22.2	25.9	22.0	19.7
6	16.7	14.4	13.5	17.3	12.3	9.7	24.3	20.9	18.7	22.0	19.4	17.9
7	11.7	9.6	7.2	14.3	10.4	8.6	24.1	21.8	20.4	19.7	17.4	17.0
8	11.9	10.5	8.8	14.8	9.5	7.0	19.8	16.7	14.4	19.2	16.6	14.5
9	10.6	8.0	5.8	14.0	8.5	5.5	23.0	21.0	18.6	22.9	20.6	18.2
10	16.0	13.1	11.6	20.8	11.6	8.5	26.5	22.8	20.1	28.0	24.1	21.3
11	17.4	14.5	13.4	20.4	14.9	12.2	29.2	25.1	21.5	26.5	23.2	20.5
12	13.2	11.1	9.2	14.8	10.5	8.7	25.2	21.3	18.5	22.5	19.6	17.1
Mean	14.0	12.5	10.3	16.8	11.4	9.0	23.5	20.3	17.9	22.3	19.4	17.3

and reconstruct the first frame, take the next $M/16$ samples and reconstruct the second frame, and repeat this until the last frame. We call this configuration PPM-2D. The second variation of PPM2 consists of taking the video generated by PPM-2D and reconstructing it again with TV3D. We call this configuration PPM3.

Tables 4 and 5 show the SNR and the structural similarity (SSIM)¹ index [37] values, respectively, for the reconstruction images obtained with PPS2, PPM2, PPM-2D, and PPM3, considering Sub-sampling ratios equal to 1/4, 1/8, and 1/16. The last lines of these tables show the average results for all videos. Overall, PPM-2D and PPM3 present the best SNR results, with PPM-2D performing best on average and in most cases (8 out of 12). Although PPM3 SSIM values are higher for almost half of the videos, SSIM results are mostly in agreement with the SNR results with PPM-2D providing (on average) better quantitative results.

Comparing the PPS2 and PPM2 reconstructions, PPM2 has better results for the Sub-sampling ratio of 1/4, with an average SNR of 14 dB for PPS2 versus 16.8 for PPM2. For higher Sub-sampling ratios, PPS2 performs better

¹SSIM values vary from ‘0’ to ‘1’, with ‘1’ corresponding to an image with a quality level that is equal to the quality of the original image and ‘0’ corresponding to extremely degraded image.

Table 5: Mean SSIM for natural videos reconstructed with PPS and PPM methods, with a Sub-sampling ratio of 1/16.

Video k	PPS2			PPM2			PPM-2D			PPM3		
	4	8	16	4	8	16	4	8	16	4	8	16
1	0.506	0.337	0.193	0.387	0.223	0.129	0.471	0.309	0.194	0.496	0.342	0.224
2	0.621	0.381	0.187	0.599	0.340	0.198	0.691	0.492	0.317	0.706	0.532	0.361
3	0.299	0.193	0.092	0.321	0.130	0.068	0.471	0.314	0.195	0.469	0.322	0.208
4	0.478	0.299	0.184	0.440	0.251	0.142	0.522	0.336	0.206	0.540	0.362	0.231
5	0.494	0.305	0.163	0.407	0.169	0.085	0.771	0.644	0.509	0.648	0.502	0.403
6	0.366	0.150	0.084	0.303	0.139	0.072	0.539	0.368	0.234	0.461	0.315	0.210
7	0.251	0.135	0.069	0.239	0.119	0.070	0.536	0.367	0.253	0.394	0.254	0.183
8	0.450	0.272	0.149	0.411	0.196	0.105	0.596	0.424	0.281	0.570	0.416	0.280
9	0.407	0.229	0.120	0.233	0.103	0.053	0.385	0.257	0.179	0.404	0.268	0.189
10	0.595	0.444	0.356	0.447	0.172	0.091	0.607	0.438	0.300	0.678	0.521	0.378
11	0.476	0.202	0.161	0.364	0.174	0.095	0.643	0.457	0.300	0.566	0.407	0.280
12	0.282	0.137	0.066	0.237	0.117	0.066	0.528	0.360	0.231	0.450	0.318	0.204
Mean	0.435	0.257	0.152	0.366	0.178	0.098	0.563	0.397	0.267	0.532	0.380	0.262

than PPM2 in most cases. However, for videos 4 and 7, PPM2 performs better. Taking a closer look at these videos, we notice that the video content is affecting the results. Video 4 has 2 frames with no visual information and 14 frames consisting of a man walking. Video 7 has lots of camera movements, including a zoom, what generates fast transitions between frames. In general, we notice for videos with a high temporal activity, PPM2 performs better than PPS2.

Figure 10 shows the results obtained for video 7. Notice that, contrary to PPS2, PPM2 generate a frame with well defined edges and low inter-frame information, but with a lot of noise. PPM3 has much better results, with images with defined edges and small amounts of noise. However, PPM-2D shows even better results than PPM3. These results are in agreement with SSIM and SNR quantitative results. Note that, as each frame is reconstructed separately, there is no content mixture among adjacent frames and even less noise. This suggests that the proposed methods obtain better results for videos with a higher content transition.

Figure 11(a) shows a 600×600 detail of the first frame of video 9. Note that, in the original frame there is a shadow that covers the man’s back. This shadow moves during the video and is not present in the last frame. Figures 11(b), (c), (d), and (e) show the frame reconstructions using PPS2, PPM2, PPM-2D and PPM3, respectively, with a 1/16 Sub-sampling ratio. For the frame reconstructed with PPS2, the back of the man appears in



(a) original, frame 1



(b) original, frame 16



(c) frame 1, PPS2, SNR=7.6



(d) frame 16, PPS2, SNR=5.3



(e) frame 1, PPM2, SNR=6.7



(f) frame 16, PPM2, SNR=7.3



(g) frame 1, PPM-2D, SNR=22.5



(h) frame 16, PPM-2D, SNR=12.1



(i) frame 1, PPM3, SNR=20.2



(j) frame 16, PPM3, SNR=11.2

Figure 10: Reconstructions of the first and last frames of video 7, using PPS2, PPM2, PPM-2D, and PPM3, at a Sub-sampling ratio of 1/16 (6.25% of samples). In these examples, the methods performed differently over the whole frames.



(a) original



(b) PPS2



(c) PPM2



(d) PPM-2D



(e) PPM3

Figure 11: Results obtained for video 9, using the acquisition methods PPS2, PPM2, PPM-2D, and PPM3, with a Sub-sampling ratio of 1/16. The white marks represent regions where the 4 methods performed very differently.

every frame (although it is not visible in the original). Therefore, PPS2 is not able to separate the content of adjacent frames, as a consequence of temporal dependency. PPM2 generates a noisy reconstruction. Both PPM-2D and PPM3 got much better results, with PPM-2D showing slightly better SNR values than PPM3. Visually, PPM3 results seem to be a little better,

showing less of the characteristic TV minimization artifacts. Although the SNR values for this video are higher for PPM2, the SSIM values are higher for PPM3 in accordance with the qualitatively results.

6. Conclusions

We proposed a new video acquisition method for high-speed reconstruction. The per-pixel mirror-based method (PPM) spatially mixes the light using micro-mirrors, in such way that no light is discarded. To maintain time independence among measurements, light is separated for each subframe. Up to our knowledge, no method in literature ensures both time independence and no information loss. Therefore, we believe PPM is a promising technique.

We performed a set of reconstruction simulations with the proposed acquisition methods and the methods available on literature. For natural images, the proposed method (PPM) showed the best results. For the synthetic sparse videos, PPM2 presented the best results in terms of reconstruction quality, showing a great SNR improvement when compared to other methods. When tested on natural high definition videos, PPM2 performs better than PPS2 for videos with a high spatial activity. The introduction of two PPM2 adaptations (PPM-2D and PPM3) improved the results, providing reconstructed video frames with a very good quality (both quantitatively and qualitatively).

The main limitation of the proposed method is the HPDMD required for its implementation, which at the moment is not commercially available. Therefore, future works include building a HPDMD prototype and implementing other variations of the PPM methodologies.

Acknowledgements

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), in part by the University of Brasília, and in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

- [1] U. Muehlmann, M. Ribo, P. Lang, A. Pinz, A new high speed cmos camera for real-time tracking applications, in: *Robotics and Automa-*

- tion, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on, Vol. 5, IEEE, 2004, pp. 5195–5200.
- [2] M. Postema, A. van Wamel, J. Folkert, N. de Jong, High-speed photography during ultrasound illustrates potential therapeutic applications of microbubbles, *Medical physics* 32 (12) (2005) 3707–3711.
 - [3] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, W. T. Freeman, The visual microphone: Passive recovery of sound from video, *ACM Trans. Graph* 33 (4) (2014) 79.
 - [4] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, M. Levoy, High performance imaging using large camera arrays, *ACM Transactions on Graphics* 24 (3) (2005) 765–776.
 - [5] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, M. Horowitz, High-speed videography using a dense camera array, in: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2, IEEE, 2004, pp. II–294.
 - [6] W. Feng, F. Zhang, X. Qu, S. Zheng, Per-pixel coded exposure for high-speed and high-resolution imaging using a digital micromirror device camera, *Sensors* 16 (3) (2016) 331.
 - [7] J. Wang, M. Gupta, A. C. Sankaranarayanan, Licens-a scalable architecture for video compressive sensing, in: *Computational Photography (ICCP), 2015 IEEE International Conference on*, IEEE, 2015, pp. 1–9.
 - [8] R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, A. K. Katsaggelos, High spatio-temporal resolution video with compressed sensing, *Optics express* 23 (12) (2015) 15992–16007.
 - [9] L. Zhang, T. Portz, Video generation with temporally-offset sampling, *uS Patent 8,958,649* (Feb. 17 2015).
 - [10] A. Veeraraghavan, D. Reddy, R. Raskar, Coded strobing photography: Compressive sensing of high speed periodic videos, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33 (4) (2011) 671–686.

- [11] J. Holloway, A. Sankaranarayanan, A. Veeraraghavan, S. Tambe, Flutter shutter video camera for compressive sensing of videos, in: Computational Photography (ICCP), 2012 IEEE International Conference on, IEEE, 2012, pp. 1–9.
- [12] External trigger modes supported by point grey cameras, <https://www.ptgrey.com/KB/10250>, accessed: 2015-07-15.
- [13] D. Reddy, A. Veeraraghavan, R. Chellappa, P2c2: Programmable pixel compressive camera for high speed imaging, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 329–336.
- [14] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, S. K. Nayar, Video from a single coded exposure photograph using a learned over-complete dictionary, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 287–294.
- [15] J. A. Lima, C. J. Miosso, M. C. Q. Farias, Per-pixel mirror-based acquisition method for video compressive sensing, in: 2014 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 1058–1062.
- [16] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *Information Theory, IEEE Transactions on* 52 (2) (2006) 489–509.
- [17] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *Information Theory, IEEE Transactions on* 52 (12) (2006) 5406–5425.
- [18] D. L. Donoho, Compressed sensing, *Information Theory, IEEE Transactions on* 52 (4) (2006) 1289–1306.
- [19] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on pure and applied mathematics* 59 (8) (2006) 1207–1223.
- [20] R. G. Baraniuk, Compressive sensing [lecture notes], *Signal Processing Magazine, IEEE* 24 (4) (2007) 118–121.

- [21] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *Information Theory, IEEE Transactions on* 52 (2) (2006) 489–509.
- [22] D. L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Transactions on Information Theory* 47 (7) (2001) 2845–2862.
- [23] C. Li, An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing, Ph.D. thesis, Citeseer (2009).
- [24] C. Li, W. Yin, Y. Zhang, Users guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms, *CAAM report* 20 (2009) 46–47.
- [25] C. Li, Compressive sensing for 3d data processing tasks: applications, models and algorithms, Ph.D. thesis, Rice University (2011).
- [26] C. Miosso, R. Von Borries, J. Pierluissi, Compressive sensing method for improved reconstruction of gradient-sparse magnetic resonance images, in: *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, IEEE, 2009, pp. 799–806.
- [27] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, R. G. Baraniuk, Compressive imaging for video representation and coding, in: *Picture Coding Symposium*, 2006.
- [28] R. Raskar, A. Agrawal, J. Tumblin, Coded exposure photography: motion deblurring using fluttered shutter, *ACM Transactions on Graphics (TOG)* 25 (3) (2006) 795–804.
- [29] A. Agrawal, M. Gupta, A. Veeraraghavan, S. G. Narasimhan, Optimal coded sampling for temporal super-resolution, in: *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on, IEEE, 2010, pp. 599–606.
- [30] M. Gupta, A. Agrawal, A. Veeraraghavan, S. G. Narasimhan, Flexible voxels for motion-aware videography, in: *Computer Vision—ECCV 2010*, Springer, 2010, pp. 100–114.

- [31] M. Douglass, DMD reliability: a MEMS success story, in: *Micromachining and Microfabrication*, International Society for Optics and Photonics, 2003, pp. 1–11.
- [32] S. K. Nayar, V. Branzoi, T. E. Boult, Programmable imaging: Towards a flexible camera, *International Journal of Computer Vision* 70 (1) (2006) 7–22.
- [33] G. Bub, M. Tecza, M. Helmes, P. Lee, P. Kohl, Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging, *Nature methods* 7 (3) (2010) 209–211.
- [34] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, R. G. Baraniuk, A new compressive imaging camera architecture using optical-domain compression, in: *Electronic Imaging 2006*, International Society for Optics and Photonics, 2006, pp. 606509–606509.
- [35] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, R. G. Baraniuk, Single-pixel imaging via compressive sampling, *Signal Processing Magazine, IEEE* 25 (2) (2008) 83–91.
- [36] L. A. Shepp, B. F. Logan, The fourier reconstruction of a head section, *IEEE Transactions on Nuclear Science* 21 (3) (1974) 21–43.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *Trans. Img. Proc.* 13 (4) (2004) 600–612. doi:10.1109/TIP.2003.819861.