PER-PIXEL MIRROR-BASED ACQUISITION METHOD FOR VIDEO COMPRESSIVE SENSING

Jonathan A. Lima, Cristiano J. Miosso, and Mylène C. Q. Farias

University of Brasília (UnB), Brasília, DF, Brazil

ABSTRACT

High-speed videos are essential to many types of scientific investigations. However, using high-speed cameras to directly acquire these videos is prohibitively expensive in many circumstances. This paper proposes a compressive sensing-based method for obtaining high-speed videos using low-speed cameras, which we call the Per-Pixel Mirror-Based Acquisition Method. The proposed technique is light efficient and generates time-independent samples. We compare the reconstruction results of the proposed technique with other techniques available in the literature in terms of signal-to-error (SER) ratios, for natural and synthetic videos. For the tested real videos, the proposed method provided an improvement in SER ranging from 3 to 28 dB, with respect to known techniques such as the flutter shutter and the per-pixel shutter. The actual improvement is higher for higher levels of sparsity in the used transformed representations and for lower used sub-sampling rates.

Index Terms— compressive sensing, computational camera, high-speed imaging, video acquisition.

1. INTRODUCTION

High-speed cameras are expensive and demand light efficiency and high bandwidths. Recently, techniques based on computational photography and compressive sensing [1, 2, 3] have been used in the design of high-speed cameras with low speed and cheap sensors. These techniques use a regular low speed camera to acquire measurements and, then, use linear combinations of these measurements to obtain high-speed video samples. Considering that the video is approximately sparse in some known domain, they reconstruct the high-speed video based on the optimization of some sparsity measure.

Unfortunately, the acquisition methods used for these techniques have some limitations. First, a large amount of light is thrown away. Second, the light information is added within the frames, what makes the measurements time-dependent. In this paper, we present a new acquisition method of linear measurements that overcomes these limitations. The proposed acquisition method uses a design of mirrors to reflect the light of some pixels into other pixels. This way, light is not thrown away and measurements can be time-independent, what leads to better reconstruction results.

This work was supported in part by the University of Brasília and in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

2. COMPRESSIVE SENSING

Compressive sensing [4, 5] is a method of acquisition and representation of signals at a rate significantly lower than the Nyquist rate. The acquisition is made from limited linear projections, a process that can be represented by the multiplication of a measurement matrix with the original signal. The measurement matrix has significantly fewer rows than the original signal. The signal is then reconstructed by a optimization process, using some measure of *sparsity*.

Let x be a signal of size N, k-sparse in the domain Ψ, and s the representation of x in the basis Ψ, i.e. $x = \Psi s$. Notice that s has only k non-zero elements, with $k \ll N$. Let y be the linear projection of x over a matrix Φ of size $M \times N$ ($M < N$), which we can represent as:

$$y = \Phi x = \Phi\Psi s = \Theta s, \tag{1}$$

where $\Theta = \Phi\Psi$ is an $M \times N$ matrix. We call Φ the measurement matrix. This is the acquisition process, where y has only M linear measurements of x.

Based on the measurements y, we want to reconstruct the signal x. Knowing x is sparse in the basis Ψ, we want to find the sparsest representation s' that satisfies $y = \Theta s'$. This is an optimization problem that can be computationally expensive. Surprisingly, the ℓ_1-norm minimization leads to a sparse solution and there are algorithms that solve this optimization problem in polynomial time.

In this paper, we use an algorithm known as Total Variation minimization (TV) [6] in which the average sum of finite differences in different dimensions are minimized. The TV algorithm used in this work is based on the optimization problem given by

$$\hat{s} = \text{armin}(||s'||_{TV}) \text{ subject to } y = \Theta s', \tag{2}$$

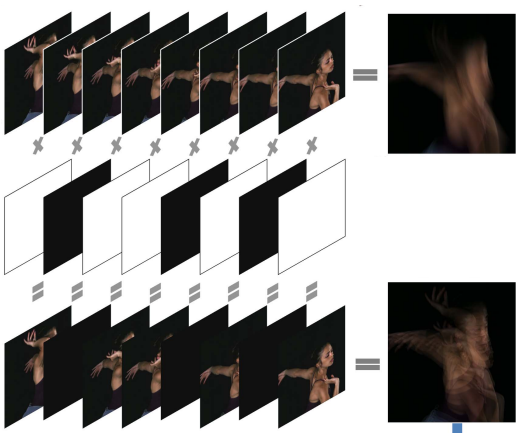


Fig. 1. Flutter Shutter acquisition method, adapted from [2]. The 1st row corresponds to the traditional acquisition in video cameras. Using the pattern of turned-on sub-frames depicted on the 2nd row, we obtain the 3rd row by making a linear combination of the sub-frames at the end of integration time.

where $\Theta = \Phi\Psi$ depends on the acquisition method, Φ represents the acquisition method, and Ψ is the transform basis. For the proposed work, we choose the TVAL3 algorithm [6]. This choice is based on the fact that, according to the extensive comparisons in [7], TVAL3 is considerably faster than the other TV minimizations and results in comparable or higher objective image qualities.

3. METHODS FOR VIDEO ACQUISITION

In this section, we describe common methods to acquire linear measurements from video cameras. Let N be the spatial resolution, FPS the Frame Per Second rate, and T the total integration time of the camera ($T = \frac{1}{FPS}$). The sub-frame is the minimum exposure time unity, corresponding to equals divisions of the frame time. k is the number of sub-frames in a frame.

Some currently available video cameras (e.g. Point Grey Dragonfly2 [1]), allow a good control of the sensor exposure. This is done by turning on and off the light flow in the array of sensors in a rate higher than the FPS of the camera. This method is known as Flutter Shutter (FS).

In the FS method, we can turn each sub-frame on and off individually. At the end of the integration time (T), we obtain an N-length image in which each pixel contains all the light corresponding to the turned-on sub-frames. This acquisition method is originally used with 67% of the sub-frames turned on [2], which does not correspond to a perfect light efficiency. An illustration of the FS acquisition scheme is presented in Figure 1.

[1] http://ww2.ptgrey.com/IEEE-1394/dragonfly-2

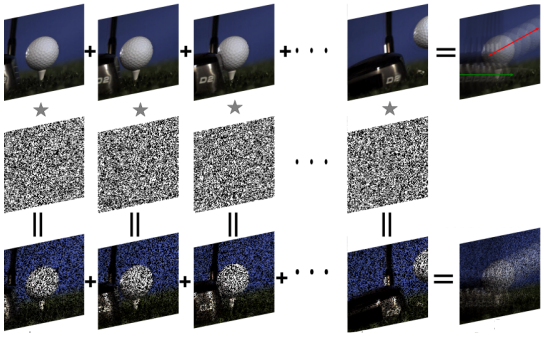


Fig. 2. PPS2 scheme, adapted from [1]. The 3 rows correspond respectively to the traditional acquisition in video cameras, the pattern of turned on and off pixels for each sub-frame, and this pattern applied to the scene. At the end of the integration time, we get a linear combination of the sub-frames with the corresponding patterns.

Another acquisition method is the Per-Pixel Shutter (PPS). In this method, we can turn on and off each one of N pixels separately for each sub-frame. This allows a greater exposure control than what is achieved by the FS method. Commercial cameras do not yet have an internal device that implements this process. So, generally, an external optical device is added to the camera optical system. The most commons devices are made with a Digital Micro-mirror Device (DMD) [8].

In the literature, the PPS method is used in 2 configurations. In the first one (PPS1), for each frame, each pixel can only be turned *on* in one of the sub-frames. This way, at the end of the frame time, each of the N pixels receives light once. Hence, each sub-frame does not have overlapping pixels, thus keeping the time independence of the sub-frames. The Flexible Voxels [9] and a dictionary learning technique proposed by Hitomi *et al.* [3] use this configuration. The disadvantage of PPS1 is that a big portion of the light information is thrown away, i.e. for k sub-frames only $(100/k)$ % of the light is kept.

In the second configuration of the PPS method (PPS2), the pixels are turned on and off independently of the other sub-frames. At the end of the integration time (T), we obtain an image that, for each pixel, contains light information of different sub-frames. The computational cameras techniques Single Pixel Camera [10] and Programmable Pixel Compressive Camera [1] use this acquisition method configuration. PPS2 has light information loss of approximately 50%. Also, there is no time-independence among the sub-frames. Figure 2 depicts the PPS2 acquisition scheme.

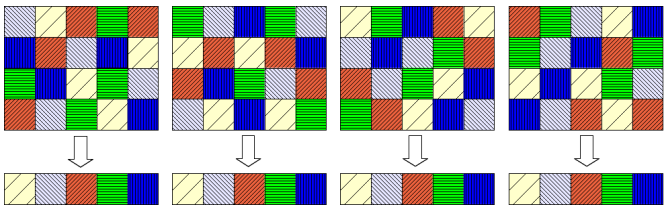


Fig. 3. PPM acquisition scheme for $k = 4$ sub-frames per frame. Each 4 by 5 matrix represents the mirror pattern of each sub-frame. The row below shows the sets of pixels chosen to receive the light. For each sub-frame, the mirrors in the matrix of a certain color redirect the light to the pixels of same color (i.e., the pixels receive the sum of the light from the mirrors of the same color).

4. PER-PIXEL MIRROR-BASED ACQUISITION METHOD

The acquisition methods presented in the previous section have two disadvantages. First, a big amount of light information is lost. Second, sub-frames are time-dependent because the light information from different sub-frames is added. To tackle these limitations, we propose a new acquisition method of linear measures called Per-Pixel Mirror-Based (PPM) acquisition method.

To implement the PPM method, we first have to propose a new device composed of micro-mirrors with a better angular precision than the mirrors in DMDs. This proposed high-precision digital micro-mirror device (HP-DMD) is composed of a micro-mirror array with one mirror for each pixel in the camera. Each mirror can redirect the light it receives to any other pixel and change its direction as fast as the sub-frame period. Notice that the HD-DMD can simulate the behavior of a DMD and, therefore, can be used for all previous acquisition methods.

In each sub-frame of PPM, we choose a set of N/k pixels to receive light from the mirrors of HP-DMD. Each pixel of this set receives light of k different random mirrors, in such a way that all mirrors redirect light. For the next sub-frame time, we choose another set of N/k pixels to receive light from the mirrors and we redirect the mirrors to this other set of pixels. We repeat this process for all k sub-frames, until we complete the N pixels at the end of the frame. The PPM scheme is shown in Figure 3.

Unlike FS and PPS, which turn off the light absorption, the proposed technique only redirects light from one set of pixels to another. In other words, this technique does not throw light away like the other techniques, but only redirects it. Notice that all k sets of pixels receive light at different sub-frame times. So, the different sets of N/k pixels are time-independent from each other.

We test two configurations of this technique. In the first, PPM1, the same pattern of random mirrors is chosen for each sub-frame. In the second, PPM2, the pattern of random mirrors changes for each sub-frame, as shown in Figure 3. For comparison, Table 1 shows the light efficiency and time-independence characteristics of the acquisition methods FS, PPS1, PPS2, and PPM.

Table 1. Light efficiency and time dependency of the acquisition techniques tested in this work.

Acquisition Method	**Light Efficiency**	**Time Independence**
FS	67%	No
PPS1	(100/k)%	Yes
PPS2	50%	No
PPM	100%	Yes

5. SIMULATION RESULTS

In our tests, we use an implementation of the optimization algorithm TVAL3 for 3D vectors (TV3D) [6]. The TV3D takes advantage of the video redundancies in time and space. To test the acquisition methods, we used synthetic and natural videos. Note that in both cases the acquisition process itself was simulated by computing the set of measurements as linear combinations of isolated pixels taken from the complete reference videos. The reconstructed videos were obtained from the simulated measures only and compared to the reference ones.

The synthetic videos are composed of a sequence of flat objects, consisting of rectangles and ellipses (phantoms). Figure 4(a) shows an example of a sub-frame of this type of video. The position, size, and intensity of some of these objects are varied to represent their movement and to illustrate possible occlusions. Four different sizes of videos were used in the tests: (100×100 pixels) × 128 frames, (100×100 pixels) × 256 frames, (200×200 pixels)× 128 frames, and (200×200 pixels) × 256 frames. We tested all acquisition methods (FS, PPS1, PPS2, PPM1, and PPM2) using 4 sub-sampling rates: 2×, 4×, 8×, and 16×, which correspond to acquiring 50%, 25%, 12.5%, and 6.25% of the samples, respectively.

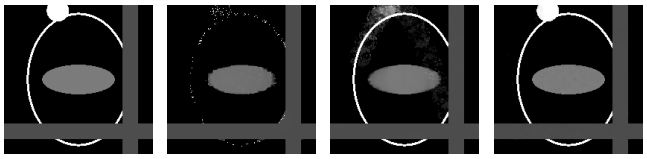
To compare the acquisition methods, we take the signal-to-error ratio (SER) between reconstructed and original sub-frames and calculate the average SER for the video. The results for the synthetic videos are shown in Table 2. PPM2 presents the best results among all tested acquisition methods, for all sub-sampling rates. In most cases, PPM1 shows the second best performance of the group. The exception corresponds to 2× sub-sampling rate, in which case the FS method has the second best performance. Nevertheless, FS shows a poor performance for sub-sampling rates higher than 2. Other exception is for PPS2, which shows good results at higher sub-sampling

Table 2. Average SER (dB): reconstructed synthetic videos.

Video Size	Acquisition Method	Sub-Sampling Rate $k=2$	$k=4$	$k=8$	$k=16$
	FS	43.5	7.4	3.5	3.4
100×	PPS1	10.5	4.8	3.0	2.4
100×	PPS2	10.5	11.1	7.8	6.6
128	PPM1	28.1	21.4	12.0	4.7
	PPM2	70.8	51.3	35.4	10.3
	FS	43.0	7.4	4.5	3.2
100×	PPS1	10.4	4.9	3.0	2.4
100×	PPS2	10.5	11.5	8.0	6.6
256	PPM1	26.2	20.7	13.2	4.7
	PPM2	77.5	51.3	33.5	9.6
	FS	38.6	9.9	6.7	5.4
200×	PPS1	12.5	12.3	3.9	2.5
200×	PPS2	12.5	14.3	10.9	9.3
128	PPM1	34.3	26.4	23.3	7.9
	PPM2	74.1	55.0	44.0	23.8
	FS	42.2	8.7	7.7	6.9
200×	PPS1	12.3	12.6	3.9	2.5
200×	PPS2	14.1	14.3	10.9	9.3
256	PPM1	37.0	31.1	24.7	9.0
	PPM2	73.5	56.9	44.9	23.2

rates. There is a great difference in performance between PPM1 and PPM2. This shows that choosing a random mirror pattern for each sub-frame makes a big difference in SER. Also, results are always better for bigger videos, but are not affected by the number of frames.

In Figures 4(b)-(d), we show the reconstructed sub-frames of the 200×200×128 video using the methods PPS1, PPS2, and PPM2 with a 16× sub-sampling rate. Notice that PPM2 does a better job at capturing the sub-frame details by better separating adjacent sub-frames. PPS2 (and FS, not shown for space limitations), on the other hand, has problems separating information from adjacent sub-frames and generates blurred frames. PPS1 is good at separating adjacent sub-frames, but it is not able to represent several details of the frame.



(a)original (b) PPS1 (c) PPS2 (d) PPM2

Fig. 4. Test performed with FS, PPS2, and PPM2 using a 200×200 ×128 video, sub-sampling rate of 16x.

In the second part of the tests, a set of 12 natural videos, obtained from The Consumer Digital Video Library (CDVL) database was used. These videos are 1280×720, 50 FPS, and 4:2:0. Given that, for the previous results the number of frames was not important, we performed our tests on 16 frames of the luminance component of these videos. For simplicity purposes, we considered the methods that performed best for synthetic videos in higher rates: PPS2 and PPM2. Additionally, we tested a second adaptation of the PPM2 in which a TV2D reconstruction output of the individual frames is used as initial solution for the TV3D optimization algorithm. We call this modification PPM3. SER results are shown in the Table 3 for sub-sampling rates of 4, 8 and 16.

Table 3. Average SER (dB): reconstructed natural videos.

Video	PPS2			PPM2			PPM3		
k=	4	8	16	4	8	16	4	8	16
1	11.3	10.0	8.4	12.7	9.1	7.3	15.8	13.2	11.5
2	15.1	13.4	11.9	17.4	12.3	10.0	20.4	17.5	15.3
3	18.4	17.9	14.6	21.7	13.9	11.1	27.1	24.0	21.7
4	12.1	10.2	9.2	15.0	12.1	10.4	17.2	14.9	13.5
5	13.3	12.0	10.1	18.9	11.8	8.4	25.9	22.0	19.6
6	16.7	14.4	13.5	17.3	12.3	9.7	22.0	19.2	17.6
7	11.7	9.6	7.2	14.3	10.4	8.6	19.7	17.8	17.5
8	11.9	10.5	8.8	14.8	9.5	7.0	19.2	16.6	14.5
9	10.6	8.0	5.8	14.0	8.5	5.5	22.9	20.6	18.2
10	16.0	13.1	11.6	20.8	11.6	8.5	28.0	24.1	21.1
11	17.4	14.5	13.4	20.4	14.9	12.2	26.5	23.3	20.5
12	13.2	11.1	9.2	14.8	10.5	8.7	22.3	19.5	16.8

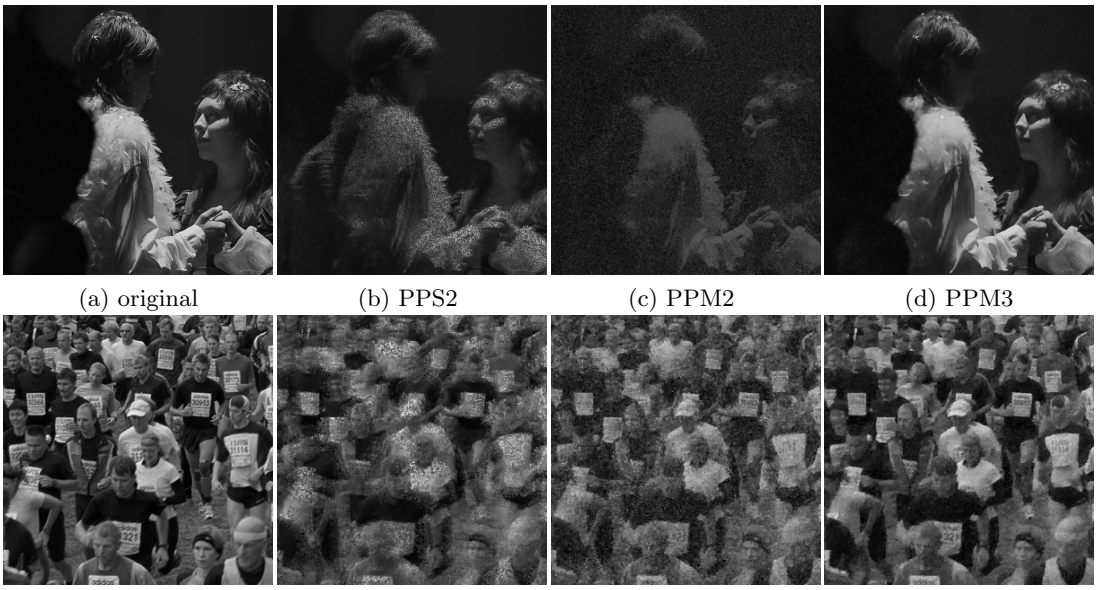
We can see in Table 3 that for the natural videos, results are a worse than what was obtained for synthetic videos. The differences are due to the fact that video sparsity in TV domain is not as good for natural videos. This decreases the performance of all methods, but PPM2 suffers a big decrease in performance. Nevertheless, the performance is better for the method PPM3.

Figure 5(a) shows a 300×300 zoom area of video 9. Notice that in the original frame, there is a shadow that is moving from behind the man. Figures 5(b), (c), and (d) show its reconstruction using PPS2, PPM2, and PPM3, respectively, with a 16× sub-sampling rate (6.25% of the samples). For the PPS2 reconstructed frame, we can see that the man's back is visible, since PPS2 did not separate the content of neighboring frames appropriately. For the PPM2 frame, the shadow appears, but the frame is noisy due to difficulties in initialization. For the PPM3 reconstructed frame, a much better result was obtained.

Figure 5(e) shows a 300×300 zoom area of video 8, which contains a crowd running and, therefore, has a lot of motion. Figures 5(f), (g), and (h) show its reconstruction using PPS2, PPM2, and PPM3, respectively, with a 4× sub-sampling rate (25% of the samples). For the PPS2 reconstructed frame, we can see that reconstruction is very blurred with information of subsequent frames. Also, the PPM2 frame is noisy. For the PPM3 reconstructed frame, we can see its details.

6. CONCLUSION

We proposed a new light-efficient method for acquiring time-independent linear measures for compressive sensing reconstruction of high-speed videos, and tested our approach by extracting simulated measurements from real and synthetic videos. For the tested videos, the pro-



(a) original (b) PPS2 (c) PPM2 (d) PPM3

(e) original (f) PPS2 (g) PPM2 (h) PPM3

Fig. 5. Test performed with PPS2, PPM2, and PPM3 for natural videos (a) v9 @ 16× sampling rate and (b) v8 @ 4× sampling rate.

posed method showed better results than the methods available in the literature, for all considered sub-sampling rates. For natural high-definition videos, we showed that the time-independence property of our technique can be used to improve the quality of reconstructed videos by including a better initialization procedure.

Future work would include a design of a reconstruction algorithm that better satisfies the compressive sensing requirements of restricted isometry property and incoherence. We also want to implement the hardware for acquiring the linear measurements required by our method, and evaluate its performance in a complete system.

7. REFERENCES

[1] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *Comp. Vision and Patt. Recog., 2011 IEEE Conf. on.* IEEE, 2011, pp. 329–336.

[2] J Holloway, AC Sankaranarayanan, A Veeraraghavan, and S Tambe, "Flutter shutter video camera for compressive sensing of videos," in *Computational Photography (ICCP), 2012 IEEE International Conf. on.* IEEE, 2012, pp. 1–9.

[3] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Computer Vision (ICCV), 2011 IEEE International Conf. on.* IEEE, 2011, pp. 287–294.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Inf. Theory, IEEE Trans.*, vol. 52, no. 2, pp. 489–509, 2006.

[5] D. Donoho, "Compressed sensing," *Information Theory, IEEE Trans.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] C. Li, W. Yin, and Y. Zhang, "User's guide for TVAL3: TV min. by augmented Lagrangian and alternating direction algorithms," *CAAM Report*, 2009.

[7] Chengbo Li, Wotao Yin, Hong Jiang, and Yin Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, 2013.

[8] M. Douglass, "DMD reliability: a MEMS success story," in *Micromachining and Microfabrication.* Int. Society for Optics and Photonics, 2003, pp. 1–11.

[9] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. Narasimhan, "Flexible voxels for motion-aware videography," in *ECCV 2010*, pp. 100–114. Springer, 2010.

[10] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Proc. Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, 2008.