An In-Depth Analysis of Compressive Sensing for High Speed Video Acquisition

Ana Serrano
Universidad de Zaragoza

Diego Gutierrez
Universidad de Zaragoza

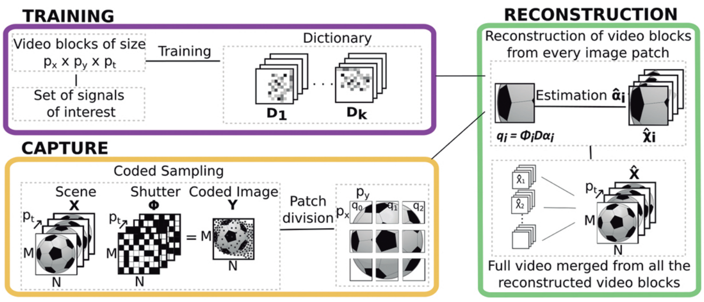
Belen Masia
Universidad de Zaragoza

Traditional video capture is limited by the trade-off between spatial and temporal resolution, and achieving both is only possible with highly specialized and expensive hardware. In this aproach, we make use of a single-shot, high-speed video capture system based on compressive sensing, in order to overcome this limitation. This allows capturing a video sequence by coding the temporal information in a single frame, and then reconstructing the full video sequence from this single coded image, relying on dictionary learning for sparse video representation. We build from the work of Liu et al. [2013], performing an in-depth analysis of the parameters of influence in the framework and providing insights for future developments of similar systems.

Framework

We overcome the bandwidth limitation of traditional video capture systems with a compressive sensing approach. In practice, this allows us to capture a video sequence by coding the temporal information within a single image. Our framework is capable of recovering a full high-speed video sequence from this single coded image.

To do this we make use of a compressive sensing framework that can be described in three main key steps.



- The training step allows us to learn fundamental building blocks (atoms) from high-speed videos, and create an overcomplete dictionary capable of sparse video representation.

$\boxed{X = D\alpha}$ ⇦ *The α coefficients represent sparsely the scene X in the domain of the dictionary D*

- In the capture step we code the temporal information with a per-pixel shutter that samples different time instants for every pixel.

$\boxed{Y = \phi X}$ ⇦ *The temporal information of the scene X is coded within the image Y with the shutter ϕ*

- We pose the reconstruction as a L_1 minimization problem, and recover video blocks as independent problems. We then combine these blocks to compose the full video sequence.

$$\boxed{Y = \phi D\alpha} \Rightarrow \boxed{\min \|\alpha\|_1 \text{ subject to } \|Y - \phi D\alpha\|_2^2 < \varepsilon}$$

Related Work

- Donoho [2006] and Candes et al. [2006] formalized the compressive sensing theory. Many works have been devoted to applying this theory in several fields.
- Liu et al. [2013] proposed a framework for high-speed video capture under a compressive sensing approach. We build on their work and provide insights on design choices for improved performance of the framework.
- Gupta et al. [2009] proposed a method to recover high spatial resolution videos from low resolution sequences and a few frames captured at a higher resolution.
- Wilburn et al. [2004] proposed an approach based in a dense camera array.

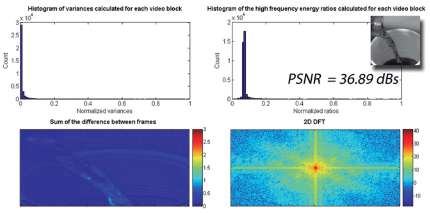
Problem & Motivation

Over the last years, high speed video capture technologies have been growing due to the neccessity of capturing information at high temporal and spatial resolutions. However, traditional cameras face a trade-off between these two resolutions, making very difficult to capture high speed videos at high spatial resolutions.
Recent advances in compressive sensing open the way to find an alternative to overcome this limitations.

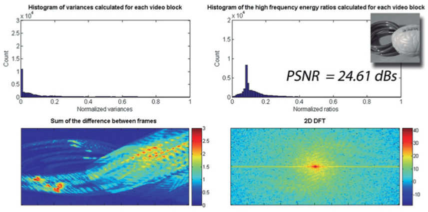
Analysis

We performed an analysis of the input scenes and observed a correlation between their characteristics and the quality (PSNR and MS-SSIM) of the final reconstructed video sequence.

- Histogram of variances calculated for each video block. *Calculate the histogram of variances of 3-D blocks, and its statistical descriptors (mean, standard deviation, skewness and kurtosis).*
- Histogram of the high frequency ratios calculated for each video block. *Add the one-dimensional DFTs along the temporal dimension of every pixel in the block. The ratio is the high frequency energy of this signal for a given threshold divided by the total energy. Calculate the hisgram of the ratios and the same statistical descriptors as above.*
- Sum of the difference between frames. *Sum the difference frame to frame and then sum all the pixels of the resulting image to obtain a single value.*
- 2-D DFT. *For a representative frame of the sequence, perform the 2-D DFT and calculate the ratio between the high spatial frequency energy for a given threshold and the total energy of the image.*

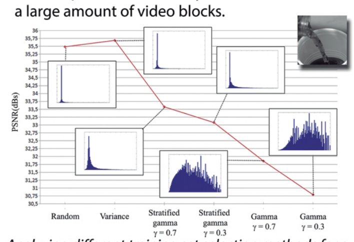


Characterization methods for two sample videos. Insets show a representative frame of each sequence and the PSNR of the reconstruction.



The standard deviation of histograms of the high frequency ratios shows a high correlation with the quality of the reconstructions ($\rho_{Pearson}$ = -0.9636, p-value = 0.0020, and $\rho_{Spearman}$ = -1, p-value = 0.0028)

We also performed an analysis of different methods for selecting a training set for dictionary learning over a large amount of video blocks.

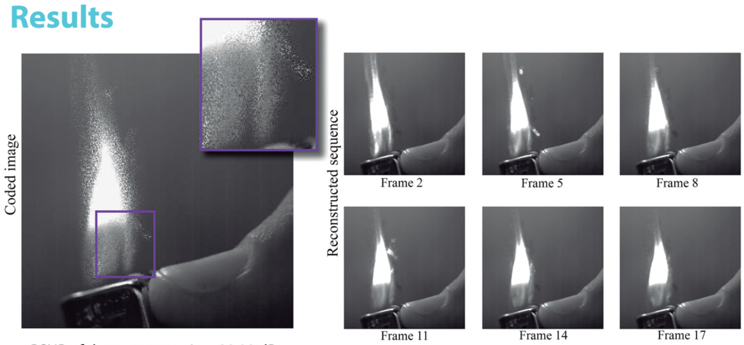


Analyzing different training set selection methods for a sample video

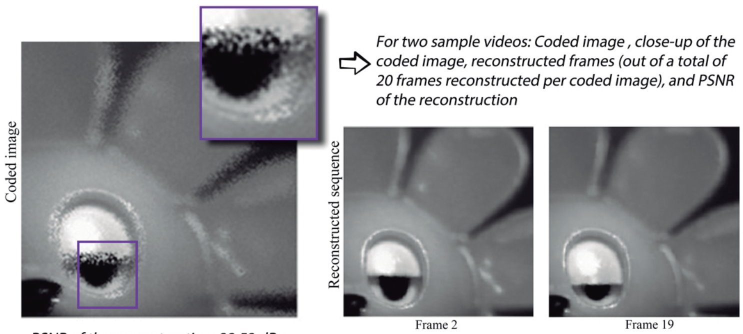
- Random sampling
- Variance sampling. *Blocks with high variance are most likely to be chosen.*
- Stratified gamma sampling. *Sort blocks by increasing variance and we perform stratified sampling under a gamma curve with different gamma values.*
- Gamma Sampling. *Choose samples from the sorted blocks following a gamma curve sampling.*

Choosing a set of training blocks with the variance method achieves better performance

Results



PSNR of the reconstruction: 29.09 dBs



PSNR of the reconstruction: 33.53 dBs

⇨ *For two sample videos: Coded image , close-up of the coded image, reconstructed frames (out of a total of 20 frames reconstructed per coded image), and PSNR of the reconstruction*

References

Candes E., Romberg J., Tao T.: Robust incertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 2006.

Donoho D. Compressed Sensing.: *IEEE Transactions on Information Theory, 2006.*

Gupta A., Bhat P., Dontcheva M., Curless B., Deussen O., Cohen M.: A Enhancing and experiencing space-time resolution with videos and stills. *IEEE International Conference on Computational Photography (ICCP)*, 2009.

Liu D., Gu J., Hitomi Y., Gupta M., Mitsunaga T., Nayar S.: Efficient Space-Time Sampling with Pixel-wise Coded Exposure for High Speed Imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.*

Wilburn B., Joshi N., Vaish V., Levoy M., Horowitz M.: High-speed videography using a dense camera array. *Computer Vision and Pattern Recognition (CVPR)*, 2004.

Graphics and Imaging Lab

ICCP 2015
International Conference on Computational Photography
April 24-26. Rice University. Houston, TX

This work is partially funded by the Spanish Ministry of Education.

Contact: anase@unizar.es
bmasia@unizar.es