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

Ana Serrano, Belen Masia and Diego Gutierrez
Universidad de Zaragoza

Compressive sensing [1, 2] is a novel technique that has been gaining increasing importance in the last years. It states that a signal can be perfectly recovered when sampled at rates below those dictated by the Shannon-Nyquist theorem provided this signal is sparse in some basis (or *dictionary*) and the sampling pattern meets some conditions. One of the applications showing the great potential of compressive sensing is the capture of high speed high resolution video in a single frame. In their work, Hitomi et al. [3, 4] present a framework capable of recovering a video sequences from a single image where the temporal information is coded with a per-pixel shutter that samples different time instants for every pixel. In this work, we have performed an in-depth analysis of the system they propose, running extensive tests over a range of parameters of influence, possible algorithms of choice, and characteristics of the input videos. We show that there is room for improvement in a number of aspects of their framework, which are detailed below.

In the *dictionary training step* we analyze the method for the selection of the training set (training patches) from a large collection of input videos. We show that the use of a stratified sampling with three levels (which may be regarded as a simplification of coresets [5], since the sampling is done with respect to patch variance), when compared to a random selection, decreases the reconstruction time an average 10% while preserving or improving the quality of the results. Figure 1 shows the quality of the reconstruction for a variety of selection methods tested. We also analyze the performance of a number of training algorithms: In particular, we show that the Online Learning algorithm [7] drastically decreases training time while yielding comparable results in terms of quality of the reconstruction.

In the *reconstruction step* we analyze alternative algorithms for solving the optimization problem. In particular we show improvements using an implementation of the LARS algorithm [8] for solving the Lasso problem over the OMP [9] algorithm proposed by the authors.

Finally, we also compute a series of statistics from a collection of videos (see Figure 2) and provide some insights about the correlation of these statistics and the quality of the reconstruction obtained for those videos, with the goal of gaining insight about the quality of the result based on the characteristics of the input video.

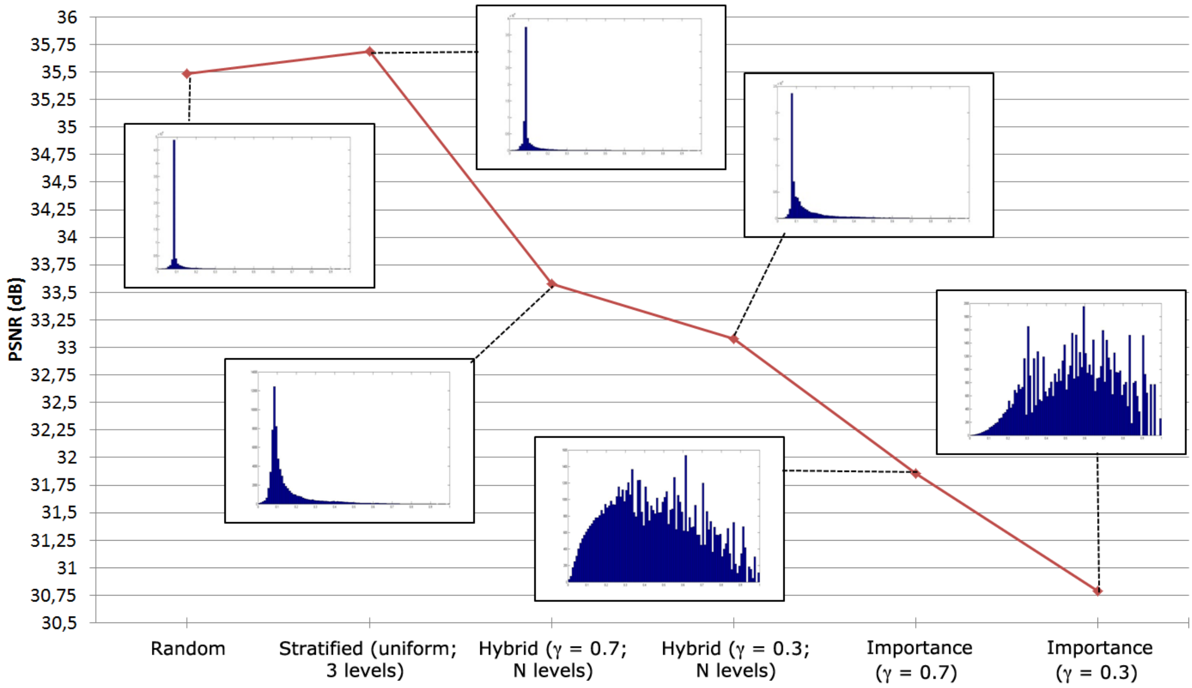

Figure 1. Analyzing different training set selection methods. The graph shows the PSNR of the reconstruction for a sample video (*PouringSoda*, a representative frame of which is shown in Figure 2) using different dictionaries. The dictionaries are trained from the same set of training videos, but the method used to select the patches for the training from those videos differs in the six cases shown, and is based on different sampling strategies of the patches depending on their variance. The best PSNR is obtained for a stratified sampling with three levels, with uniform sampling between levels. Insets show the resulting histogram of variances of the sampled patches used to train each dictionary according to each particular method.

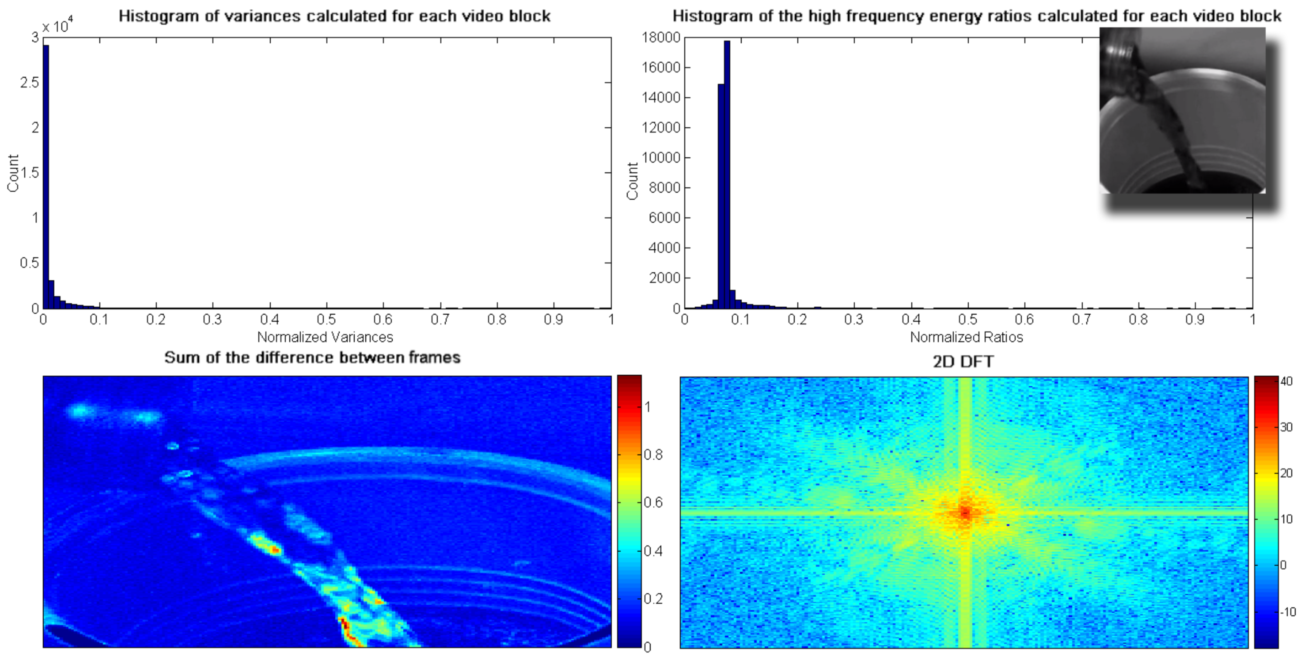

Figure 2. Characterization of a sample video (*PouringSoda*, a representative frame of which is shown in the top right inset). *Top row*: Histogram of the variances computed for every video block (*left*) and histogram of the high frequency energy density ratios[1] calculated for every video block (*right*). *Bottom row*: Sum of the difference between all consecutive frames of the video (*left*) and two dimensional DFT (*right*). In particular, the standard deviation of the histogram of high frequency energy ratios has shown a high correlation with the quality of the reconstruction results expressed in terms of the PSNR ($\rho_{Pearson}$ = -0.9636 with an associated p-value = 0.0020, and $\rho_{Spearman}$ = -1 with an associated p-value = 0.0028).

[1] The high frequency energy density ratio of a video block is calculated by performing a unidimensional DFT along the temporal dimension for every pixel of the block. Then for each block, the DFTs corresponding to all its pixels are summated and the ratio is calculated as the high frequency energy of the signal defined for a given threshold divided by the total energy.

REFERENCES

[1] Candès, E., Romberg, J., and Tao, T. 2006.: Robust incertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory,* 52, 2, 489-509.

[2] Donoho, D. L. 2006.: Compressed sensing. *IEEE Transactions on Information Theory,* 52, 4, 1289-1306.

[3] Hitomi Y., Gu, J., Gupta, M., Mitsunaga, T. , and Nayar, S. 2011.: Video from a Single Coded Exposure Photograph using a learned Over-Complete Dictionary. *IEEE International Conference on Computer Vision.*

[4] Liu D., Gu, J., Hitomi, Y., Gupta, M., Mitsunaga, T., and Nayar, S. 2013.: Efficiente Space-Time Sampling with Pixel-wise Coded Exposure for High Speed Imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99.

[5] Feigin, M., Feldman, D., and Sochen, N. 2011.: From high definition image to low space optimization. Conference on Scale Space and Variational Methods in Computer Vision, 459-470.

[6] Aharon, M., Elad, M., and Bruckstein, A. 2006.: The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation. IEEE Transactions on Signal Processing, 54, 11, 4311-4322.

[7] Mairal, J., Ponce, J., and Sapiro, G. 2010.: Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine learning Research, 11, 19-60.

[8] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. 2004.: Least angle regression. Annals of statistics, 32(2), 407-499.

[9] Tropp, J. A., and Gilbert, A. C. 2007.: Signal Recovery from Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory, 53, 12, 4655-4666.