UNIVERSITY OF CAMBRIDGE

Compressive Sensing in Video Reconstruction

Brian Azizi

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

Supervised by:
Dr Anita Faul

Department of Physics
Cavendish Laboratory
JJ Thomson Avenue
CB3 0HE
Cambridge, UK

July 26, 2016

Contents

Chapter 1

Introduction

There are three parts: A signal processing framework called *Compressive Sensing (CS)*, a pre-processing step in form of a basis transformation based on discrete wavelet transforms and a Machine Learning algorithm called *Sparse Bayesian Learning.*

Background

Chapter 2

Compressive Sensing

This section is based on [1]. The problem to be solved can be formulated as follows: Let $\boldsymbol{x} \in \mathbb{R}^N$ be a signal of interest. We do not measure $\boldsymbol{x}$ directly and it is thus unknown. Instead, we have a measurement $\boldsymbol{y} \in \mathbb{R}^M$, with $M << N$, from which we want to reconstruct $\boldsymbol{x}$. The signals $\boldsymbol{x}$ and $\boldsymbol{y}$ are related as follows:

$$\boldsymbol{\Omega x} = \boldsymbol{y} \tag{2.1}$$

where $\boldsymbol{\Omega}$ is a known $M \times N$ matrix referred to as the *sensing matrix*.

For example, in [1], the signal of interest $\boldsymbol{x}$ is an image, so that N is equal to the total number of pixels in the image and x_i is equal to the intensity of the corresponding pixel. However, we imagine that we have only access to a corrupted version of $\boldsymbol{x}$ in which random pixel values have been deleted. This is our measurement $\boldsymbol{y}$. See Figure 2 for an example. The sensing matrix $\boldsymbol{\Omega}$ corresponding to this scenario is obtained by taking the $N \times N$ identity matrix and deleting the rows that correspond to the missing entries in $\boldsymbol{x}$.

Compressive Sensing (CS) is a collection of signal processing techniques that

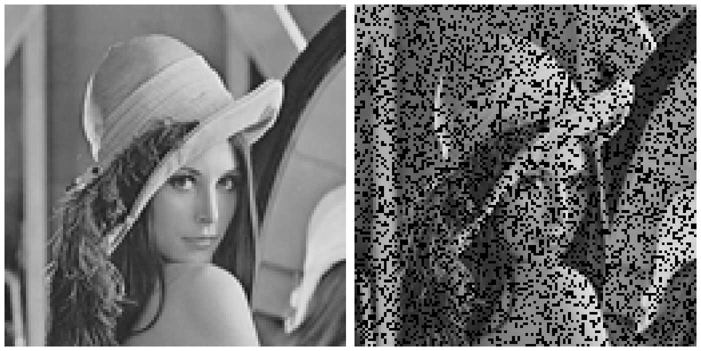

Figure 2.1: Example of a signal pair $\boldsymbol{x}$ (left) and $\boldsymbol{y}$ (right). We wish to reconstruct $\boldsymbol{x}$ from $\boldsymbol{y}$.

allow for efficient *reconstruction* (and indeed *aquisition*) of such signals by solving the underdetermined system (2.1).

Of course, there are infinitely many solutions to an underdetermined system. In the CS framework, we seek to find a solution $\hat{\boldsymbol{x}}$ that is *sparsest in some domain*. By that, we mean that we want to find $\hat{\boldsymbol{x}}$ that satisfies (2.1), such that there exists a basis transformation of $\hat{\boldsymbol{x}}$ in which it has the smallest number of nonzero entries.

More concretely, we assume there exists a domain in which the desired signal $\boldsymbol{x}$ is sparse. I.e. there exists a $N \times N$ basis matrix Ψ such that $\boldsymbol{x} = \boldsymbol{\Psi w}$ and $\boldsymbol{w}$ is sparse.

The CS problem can then be expressed as follows:

$$\min ||\boldsymbol{w}||_0 \qquad \text{subject to} \qquad \boldsymbol{\Omega \Psi w} = \boldsymbol{y} \tag{2.2}$$

where $||.||$ denotes the l_0 norm, i.e. the number of nonzero components.

For a more detailed review of the CS framework, see [2].

Chapter 3

Wavelets

In this chapter, we will introduce wavelet functions and the *Discrete Wavelet Transform* (DWT). The DWT allows us to transform our signal into a new basis in which it is sparse.

$$\boldsymbol{x} = \boldsymbol{\Psi} \boldsymbol{w} \tag{3.1}$$

that takes the dense signal $\boldsymbol{x}$ and sends it into a domain in which its transformation $\boldsymbol{w} = \boldsymbol{\Psi}^T \boldsymbol{x}$ is sparse.

3.1 Discrete Cosine Transform

An aside on the DCT. Used in old JPEG standard (ref). Formulae. Interpretations. Pictures. However, before we discuss wavelets, we will briefly introduce the *Discrete Cosine Transform* (DCT)

3.2 Wavelet transforms

3.2.1 Haar wavelets

Finding a set of basis functions $\boldsymbol{\Psi}$ that achieve such a transformation lies at the heart of many lossy compression techniques. For instance, in image processing the JPEG 2000 standard is a widely used lossy compression technique that relies on this principle. The original image $\boldsymbol{x}$ is transformed into $\boldsymbol{w}$ using the so-called *Discrete Cosine Transform.* The basis matrix $\boldsymbol{\Psi}$ is orthogonal, so $\boldsymbol{x}$ and $\boldsymbol{w}$ have the same l_2 norm. In the original signal $\boldsymbol{x}$ the length is spread across many of its coefficients. On the other hand, most of the length of $\boldsymbol{w}$ is concentrated in a few of its coefficients. A large fraction of the entries in $\boldsymbol{w}$ are very close to zero. By deleting these entries in $\boldsymbol{w}$ and only storing the non-zero coefficients (and the corresponding basis functions), we can obtain a compressed version $\hat{\boldsymbol{w}}$. This allows us to significantly reduce the amount of data that needs to be stored without affecting the visual quality in the reconstructed image $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}\hat{\boldsymbol{w}}$.

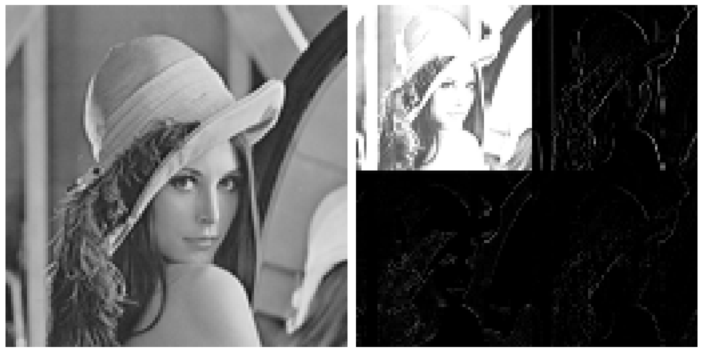

Figure 3.1: Original image $\boldsymbol{x}$ (left) and its Haar basis transformation $\boldsymbol{w}$ (right). See next chapter for more details on Haar wavelets.

It is important to note here that the choice of basis functions $\boldsymbol{\Psi}$ typically has a significant effect on the performance of the reconstruction algorithms.

The simplest wavelet basis transformation is based on the Haar wavelets. Figure 3.2.1 vizualises a basis transformation of the original image $\boldsymbol{x}$ to $\boldsymbol{w}$. This example uses a Haar wavelet basis at the first scale. We will explain the Haar wavelet transformations of images and videos in more detail in the next chapter. Dark areas correspond to small coefficients. Note that most entries in $\boldsymbol{w}$ are near zero. In practice, we approximate these entries as zero and treat $\boldsymbol{w}$ as sparse.

3.2.2 Daubechies Wavelets

Intuition. Where do the coeffs come from. Matrices. Boundary conditions.

3.3 Forming the basis matrix

1D, 2D, 3D case. Different scales

For a deeper introduction into wavelets see [3]. For more information on wavelet compression techniques, see [4].

Chapter 4

Design of the Multi-Scale Cascade of Estimations Algorithm

Bringing all building blocks together. Description and explanation of the algorithm

4.1 Interpolator

We use a sensing matrix $\mathbf{\Omega}$ that acts as signal mask. That is, we obtain the $N \times M$ matrix $\mathbf{\Omega}$ by taking the $M \times M$ identity matrix $\boldsymbol{I}_M$ and deleting $(M-N)$ rows. This corresponds to a subsampled signal in which we only measured N pixel values. For this specific class of sensing matrices, the problem of reconstructing the original signal is also known as *interpolation.*

Chapter 5

Implementation Details and Code optimization

This chapter gives a brief description of the current state of our implementation of the 3D signal reconstructer.

5.1 Haar basis functions

The RVM takes as input a target vector ($\boldsymbol{y}$) and a basis matrix ($\boldsymbol{\Psi}$). In this respect, it is agnostic about whether the signal is an image or video or of some other type alltogether. Most of this information is encoded in the basis matrix $\boldsymbol{\Psi}$. It is therefore important, and often challenging, to select a good set of basis functions.

Our current implementation uses 3-dimensional Haar wavelet basis functions. I will show how the basis matrix $\boldsymbol{\Psi}$ is constructed by briefly describing how the discrete Haar wavelet transform is performed on 1D, 2D and finally on 3D signals.

5.1.1 1D Haar wavelet transform

Consider a 1-dimensional signal $\boldsymbol{s} = \{s_1, \ldots, s_r\} \in \mathbb{R}^r$ (r for "rows"), where, for simplicity, we assume that r is a power of 2. The Haar wavelet transform can be performed at various resolution scales. The transform at the first scale is given by:

$$\boldsymbol{s} = \{s_1, \ldots, s_r\} \rightarrow \frac{1}{\sqrt{2}}\{s_1+s_2, s_3+s_4, \ldots, s_{r-1}+s_r, s_1-s_2, \ldots, s_{r-1}-s_r\} = \hat{\boldsymbol{s}}^{(1)}$$

The first half of the signal is replaced by scaled averages of adjacent elements and the second half is replaced by scaled differences of adjacent elements. By performing this transform again on the first half of $\hat{\boldsymbol{s}}^{(1)}$ while keeping the second half fixed, we get the Haar wavelet transform at the second scale $\hat{\boldsymbol{s}}^{(2)}$. To get the third scale transform $\hat{\boldsymbol{s}}^{(3)}$, we perform the initial transform on the first quarter of $\hat{\boldsymbol{s}}^{(2)}$ while keeping the rest of the signal fixed. We may continue this process until we reach the ith scale, where $2^i = r$.

From here on, we will only consider the first scale transform $\hat{\boldsymbol{s}}^{(1)}$ and we will omit the (1) superscript. We can express the transform as a multiplication by an orthogonal $r \times r$ matrix W given by

$$W = \begin{bmatrix} \Phi_r \\ \Psi_r \end{bmatrix} \tag{5.1}$$

where Φ_r and Ψ_r are $(r/2) \times r$ matrices[1] given by

$$\Phi_r = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

and

$$\Psi_r = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}$$

In the signal processing literature, Φ_r is referred to as a low pass filter, while Ψ_r is referred to as a high pass filter. Φ_r outputs an average of the signal and Ψ_r outputs the details of the signal.

5.1.2 2D Haar wavelet transform

Let $A \in \mathbb{R}^{r \times c}$ be a 2-dimensional signal (e.g. an image). For simplicity, we will assume that both r and c are powers of 2 (though not necessarily equal).

It is simple to obtain A's Haar wavelet transform $\hat{A}$ at the first scale. This is done by first applying the 1-dimensional transform individually to each column of A to obtain a temporary matrix $\hat{A}_{temp}$. Next, we apply the 1-dimensional haar wavelet transform individually to each row of $\hat{A}_{temp}$ to obtain $\hat{A}$.

We can again express the transform as a multiplication of matrices:

$$\hat{A} = \begin{bmatrix} \Phi_r \\ \Psi_r \end{bmatrix} A \begin{bmatrix} \Phi_c^T & \Psi_c^T \end{bmatrix} \tag{5.2}$$

where Φ_r and Ψ_r are as before and Φ_c and Ψ_c are of similar form but each have dimensions $(c/2) \times c$. This is the transform that was used to generate the RHS of Figure 3.2.1. We note that the high-pass filters essentially detect edges of various orientations in the image.

[1] Note that the matrix Ψ_r used here is different to the matrix Ψ that was used in the previous chapter (which corresponds to W^T here).

However, as it currently stands, we cannot use this form of the basis transformation for the reconstruction algorithm. Recall that the RVM requires a *vector* of measurements as opposed to a matrix and also that it requires a single basis matrix, not a basis transform as given in (5.2).

To do this, we store the 2-dimensional signal A as a long column vector $\boldsymbol{a}$ of length rc by pasting the individual columns of A one after another. The basis transformation of $\boldsymbol{a}$ can then be expressed as

$$\hat{\boldsymbol{a}} = W\boldsymbol{a}$$

where W is a $rc \times rc$ matrix given by

$$W = \begin{bmatrix} \Phi_c \otimes \Phi_r \\ \Phi_c \otimes \Psi_r \\ \Psi_c \otimes \Phi_r \\ \Psi_c \otimes \Psi_r \end{bmatrix}$$

The symbol $\otimes$ denotes the *Kronecker product.* The kronecker product $P \otimes Q$ between matrices P and Q with dimensions $m_P \times n_P$ and $m_Q \times n_Q$, respectively, is defined to be the block matrix

$$\begin{bmatrix} p_{1,1}Q & p_{1,2}Q & \cdots & p_{1,n_P}Q \\ p_{2,1}Q & p_{2,2}Q & \cdots & p_{2,n_P}Q \\ \vdots & \vdots & \ddots & \vdots \\ p_{m_P,1}Q & p_{m_P,2}Q & \cdots & p_{m_P,n_P}Q \end{bmatrix}$$

of size $m_P m_Q \times n_P n_Q$.

5.1.3 3D Haar wavelet transform

Let $V \in \mathbb{R}^{r \times c \times s}$ be a 3-dimensional signal such as a video. V has r rows, c columns and s slices, and we assume that r, c and s are all powers of 2. We may visualize V as a "volume" with 2 spacial dimensions and one time dimension corresponding to frames of the video.

To obtain the Haar wavelet transform $\hat{V}$ of V, we first perform the 1-dimensional transform individually on each column in every slice of V to get $\hat{V}_{temp1}$. We then perform the 1D transform on every row in every slice of $\hat{V}_{temp_1}$ to get $\hat{V}_{temp_2}$. Finally, we perform the 1D transform across the slices for every row and column to get $\hat{V}$.

However, like in the 2-dimensional case, we need to be able to pass a single vector of coefficients and a single basis matrix to the RVM. To do this, we vectorize V as follows. First, we vectorize each individual slice of V as before in the 2D case. Then, we stack all these vectors on top each other to get one very long column vector $\boldsymbol{v}$ of length rcs. The Haar wavelet transform is given by

$$\hat{\boldsymbol{v}} = W\boldsymbol{v}$$

where

$$W = \begin{bmatrix} \Phi_s \otimes \Phi_c \otimes \Phi_r \\ \Phi_s \otimes \Phi_c \otimes \Psi_r \\ \Phi_s \otimes \Psi_c \otimes \Phi_r \\ \Phi_s \otimes \Psi_c \otimes \Psi_r \\ \Psi_s \otimes \Phi_c \otimes \Phi_r \\ \Psi_s \otimes \Phi_c \otimes \Psi_r \\ \Psi_s \otimes \Psi_c \otimes \Phi_r \\ \Psi_s \otimes \Psi_c \otimes \Psi_r \end{bmatrix}$$

Comparing notation to the previous chapter, what we refer to as W here is the transpose of what was previously denoted as Ψ. And since $\boldsymbol{v} = W^T \hat{\boldsymbol{v}}$, we see that $\boldsymbol{v}$ corresponds to what was previously called $\boldsymbol{x}$.

Chapter 6

Results

We have obtained some results with our current implementation. The implementation uses the Haar wavelet transform at the first scale.

Our example video has a resolution of 128 by 128 pixels and consists of a total of 64 frames. Thus, $r = 128$, $c = 128$ and $s = 64$. Note that even for such a relatively small sample, the size of the basis matrix Ψ is $(128 * 128 * 64) \times (128 * 128 * 64) = 1048576 \times 1048576$. Even in single precision, storing this matrix would require around 4 terrabytes.

For this reason, we have split the original input signal into $8 \times 8 \times 8$ blocks and perform the algorithm on the individual blocks.

In Figures 3.1 and 3.2, we have included a sample frame from the corrupted test video and the same frame after reconstruction.

In Figure 3.1, we corrupted the video by deleting 30% of the pixel values in the first frame and deleting the same pixel values in each subsequent frame (so the same pixels are missing in each frame). Figure 3.2 uses the same corruption scheme but we deleted 50% rather than 30% of pixel values.

These initial results are promising, though clearly there are still improvements to be made.

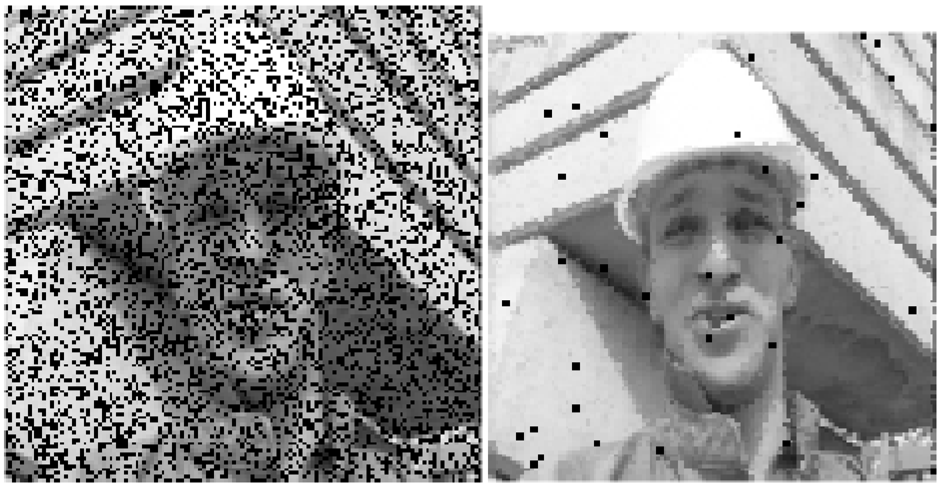

Figure 6.1: Sample frame from corrupted video (left) and the reconstructed video (right)

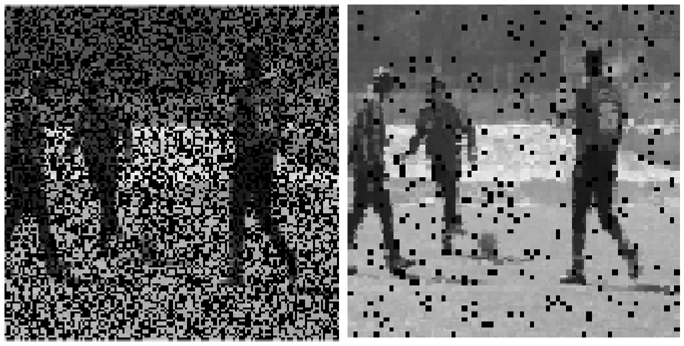

Figure 6.2: Sample frame from corrupted video (left) and the reconstructed video (right)

Chapter 7

Conclusion

Bibliography

[1] G. Pilikos, Signal reconstruction using compressive sensing, MPhil thesis, University of Cambridge (2014).

[2] E. J. Candès, M. B. Wakin, An introduction to compressive sampling, Signal Processing Magazine, IEEE 25 (2) (2008) 21–30.

[3] E. J. Stollnitz, T. D. DeRose, D. H. Salesin, Wavelets for computer graphics: a primer. 1, Computer Graphics and Applications, IEEE 15 (3) (1995) 76–84.

[4] R. A. DeVore, B. Jawerth, B. J. Lucier, Image compression through wavelet transform coding, Information Theory, IEEE Transactions on 38 (2) (1992) 719–746.