Image representation and compression via sparse solutions of systems of linear equations

Alfredo Nava-Tudela
Norbert Wiener Center, University of Maryland, College Park

Introduction

Cn u rd ths?

If you answered yes to the above question, then you have grasped what we are trying to do here, but for images. In the example above, we have compressed the sentence “can you read this?” to “cn u rd ths?,” which amounts to a reduction of six characters, 33% fewer characters than in the original sentence, but without compromising its meaning.

We can do something similar for images by way of the following algebraic trick. Suppose that you have the system of linear equations

$$\mathbf{A}\,\mathbf{x} = \mathbf{b},$$

where **A** is a full-rank matrix with more columns than it has rows, and **x** and **b** are vectors that are compatible with the matrix-vector product shown above.

This basically means that if **A** has m rows and n columns, then $m < n$, **x** is a vector with n entries, and **b** is a vector with m entries. Moreover, if we are given any such vector **b** then we can always find a solution vector **x** to the equation **A x = b** (this is what **A** being full-rank means.) In picture form this looks like this

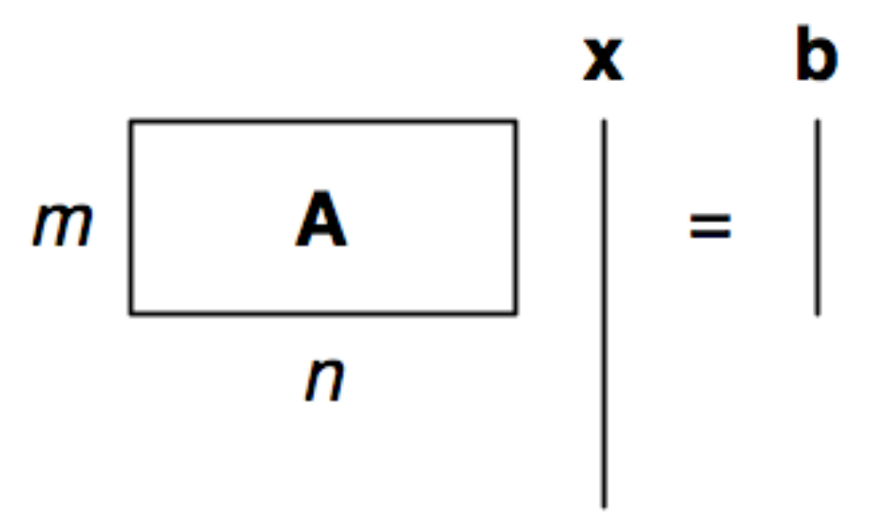

Figure 1. Underdetermined system of linear equations with infinite number of solution vectors **x** for any given “signal vector” **b**.

It is a fact that there are an infinite number of solutions to equations of the type depicted in **Fig. 1**, provided **A** is full-rank. This is what we can exploit to compress an image I. Suppose that we can somehow convert I into a vector **b** and that for some ad hoc matrix **A** we can find a vector $\mathbf{x}_0$ such that the number of nonzero entries of $\mathbf{x}_0$, from now on written as $\|\mathbf{x}_0\|_0$, is a lot smaller than the number of nonzero entries of vector **b**, $\|\mathbf{x}_0\|_0 < \|\mathbf{b}\|_0$ in our new notation. Then if we store or transmit $\mathbf{x}_0$ instead of **b** we would have compressed image I.

Suppose now that we look at a slightly different but related problem, we look for the solution of

$$(P_0^\varepsilon)\text{: } \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{A}\,\mathbf{x} - \mathbf{b}\|_2 < \varepsilon.$$

If we set $\varepsilon = 0$, we are back to the previous **A x = b** problem. However, if we allow $\varepsilon > 0$ then we have a trade-off between how close the solution $\mathbf{x}_0$ of (P_0^ε) is to **b** via $\mathbf{A}\mathbf{x}_0$, and how small $\|\mathbf{x}\|_0$ is. This is what we explore here.

Methodology

Given an 8-bit, grayscale image of size 512 by 512 pixels, we subdivide it into non-overlapping sub-images ☐ of 8 by 8 pixels. Each sub-image is converted to $\mathbf{b} = c(☐)$, an input vector in problem (P_0^ε) and, given a matrix **A** and tolerance $\varepsilon > 0$, we obtain a reconstruction of the original image by putting together in the right order $☒ = c^{-1}(\mathbf{A}\,\mathbf{x}_0)$, where $\mathbf{x}_0$ is the solution to (P_0^ε) obtained by Orthogonal Matching Pursuit (OMP), a greedy algorithm. OMP guarantees that

$$\| c(☒) - c(☐) \|_2 < \varepsilon$$

We do this for multiple values of $\varepsilon > 0$ to study the trade-off between error and compression.

Results

We obtain better compression for a given tolerance $\varepsilon > 0$ when we form matrix **A** by combining two bases (**Fig. 3**) as opposed to using only one, such as in the JPEG or JPEG 2000 standards. This can be seen in **Fig. 4** where the compression vs tolerance graphs for image *Barbara* (**Fig. 2.a**), are drawn. The reconstructions for the standard OMP implementation with $\varepsilon = 32$ (**Fig. 2.b1**) and the modified OMP with $\delta = 0.94$ (**Fig. 2.c1**) have, peak signal-to-noise ratios (PSNR) of 36.9952 dB and 32.1482 dB, respectively. In **Figs. 2.b2** and **2.c2** the difference of the original and the reconstruction is quantified by the structural similarity index (SSIM).

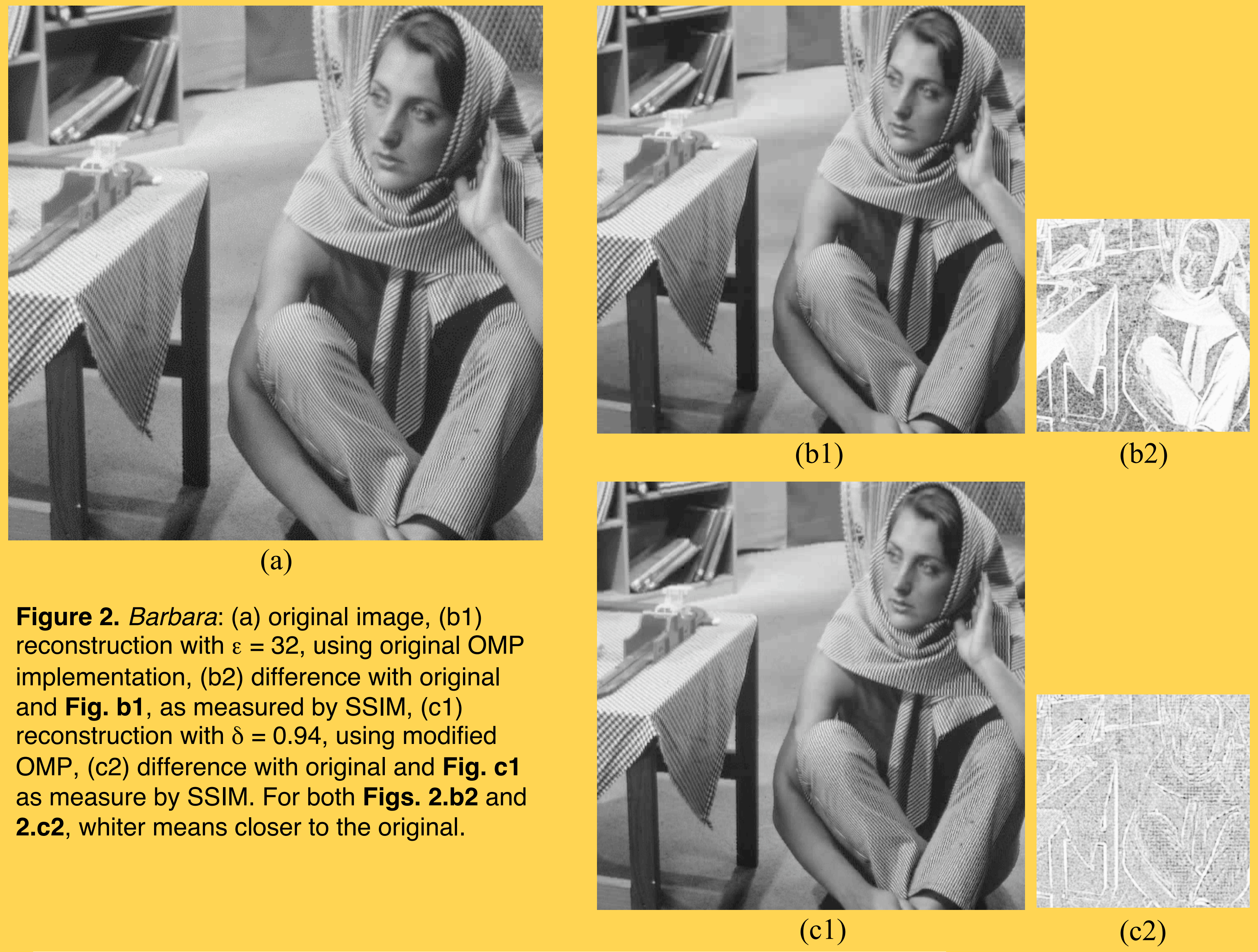

Figure 2. *Barbara*: (a) original image, (b1) reconstruction with $\varepsilon = 32$, using original OMP implementation, (b2) difference with original and **Fig. b1**, as measured by SSIM, (c1) reconstruction with $\delta = 0.94$, using modified OMP, (c2) difference with original and **Fig. c1** as measure by SSIM. For both **Figs. 2.b2** and **2.c2**, whiter means closer to the original.

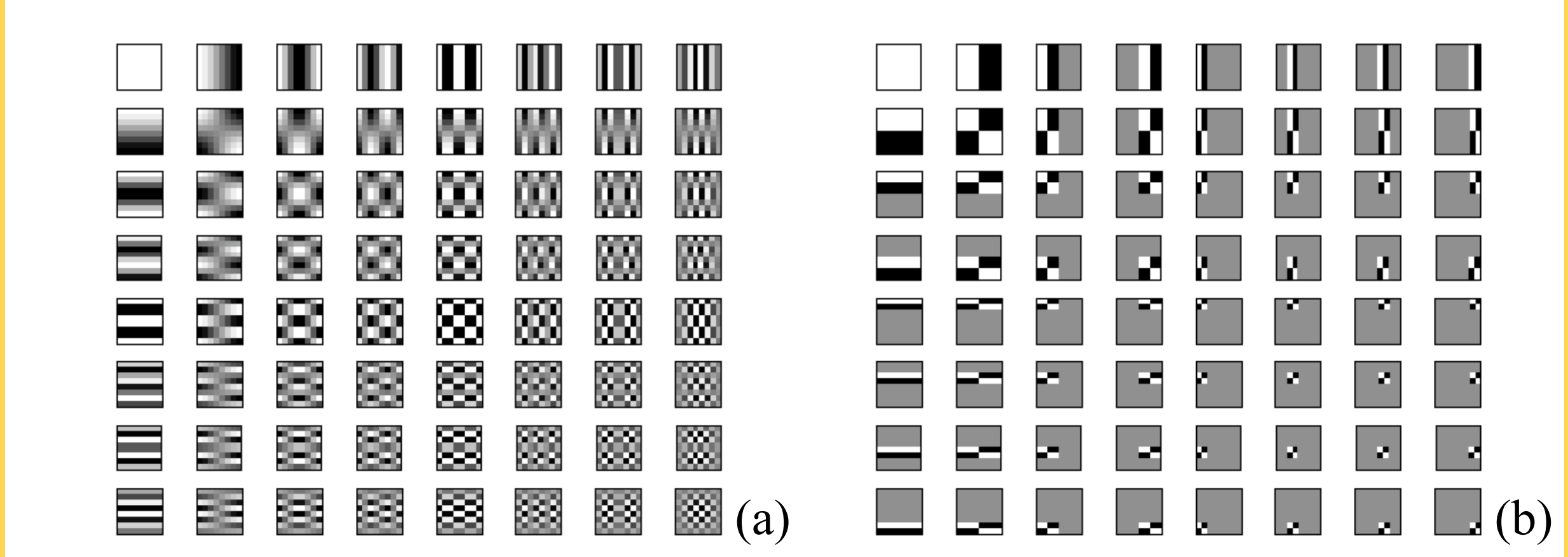

Figure 3. The 2-dimensional basis elements used to build (a) basis $\mathrm{DCT}_{2,j}$ and (b) basis $\mathrm{Haar}_{2,j}$. One of the two classes of matrices $\mathbf{A} = [\mathrm{DCT}_{2,j}\ \mathrm{Haar}_{2,j}]$ that we used in problem (P_0^ε) concatenate both bases.

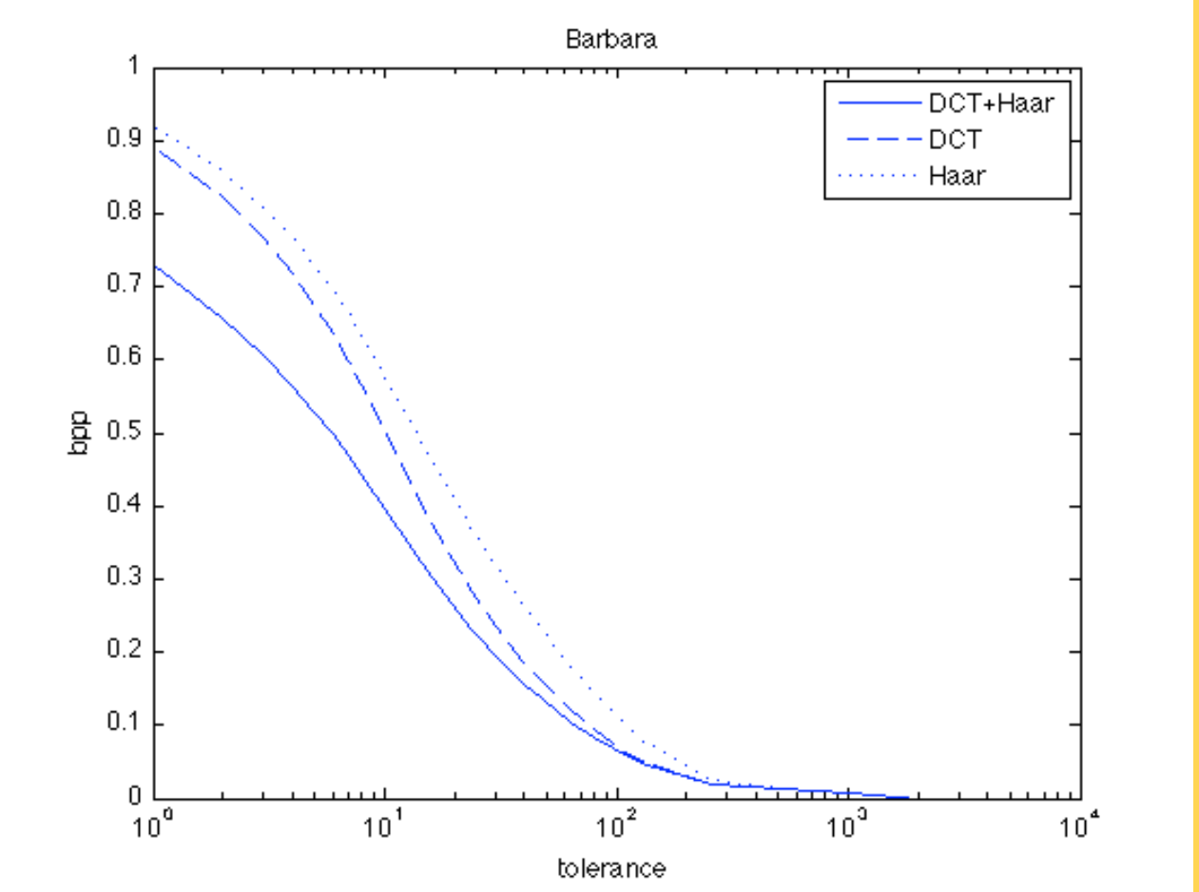

Figure 4. Normalized bit-rate in bits per pixel (bpp) vs tolerance ε. The dashed line corresponds to the DCT basis (**Fig. 3.a**), the dotted line shows the Haar basis (**Fig. 3.b**), and the solid line the combination of both bases. Lower bpp values mean more compression.

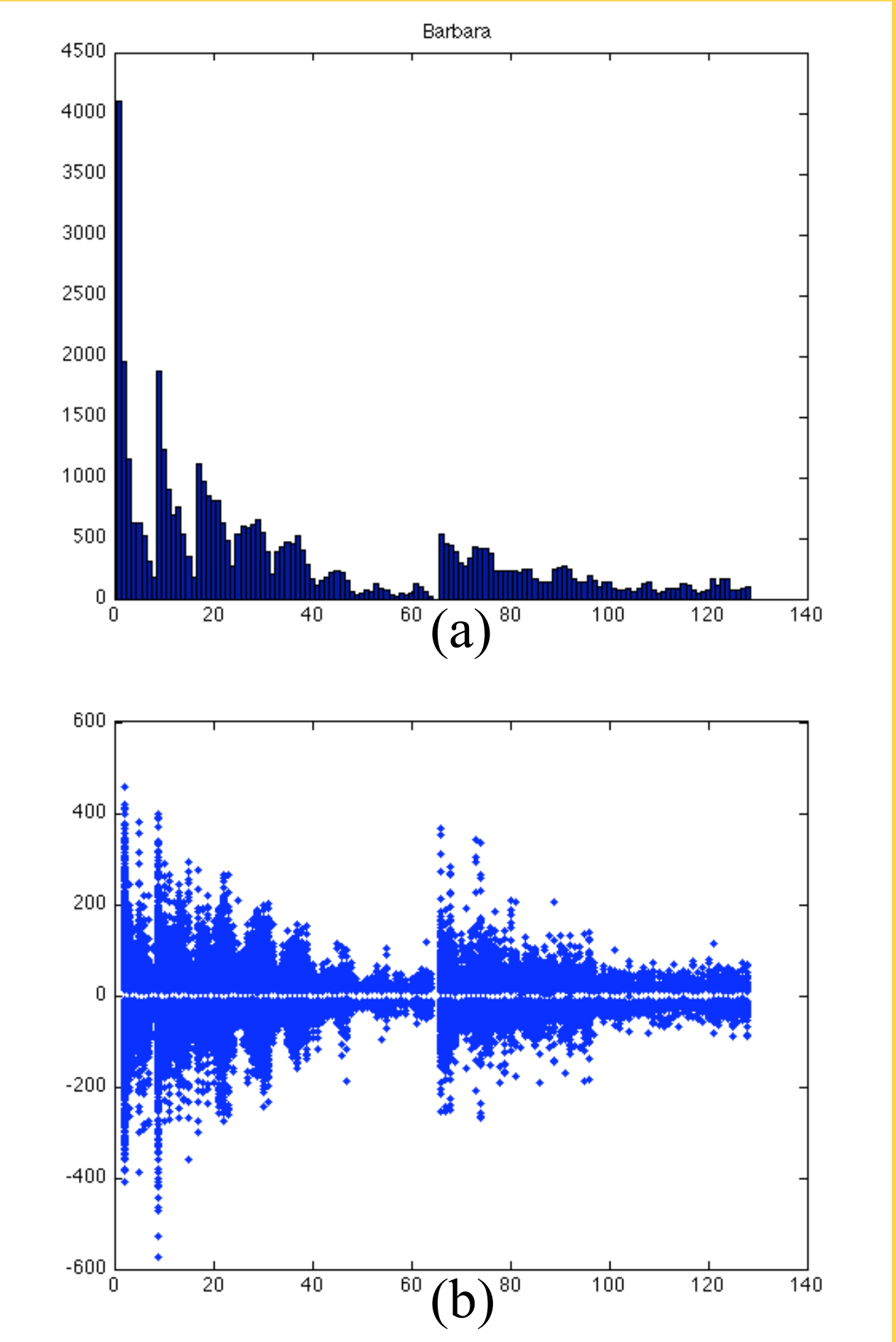

Figure 5. Histogram and coefficient distribution for a tolerance value of $\varepsilon = 32$: (a) the histogram that shows the number of times a given column of $\mathbf{A} = [\mathrm{DCT}_{2,3}\ \mathrm{Haar}_{2,3}]$ has been chosen, (b) the distribution of the coefficients that multiply the chosen columns.

Conclusions

The ever-increasing volume of images in use by many multimedia channels, like Internet web sites or e-books, to name a few, demands novel techniques to represent and compress visual information in order to reduce the strains on the limited channel capacity of mobile communication devices, or the storage requirements for this type of information.

We have explored the merits of utilizing approximate sparse solutions of systems of linear equations (**A x = b**) for this purpose and we have found that we can achieve better compression of a given “signal” vector **b** by combining two bases (stored in matrix **A**) rather than using only one to represent **b** as is done in common compression standards JPEG and JPEG 2000.

We have also modified the standard OMP procedure to find solutions of (P_0^ε) incorporating image quality indicators that take into account the peculiarities of the human visual system. This resulted in improved compression without perceptible visual degradation as measured by the structural similarity index (SSIM), and the mean structural similarity index (MSSIM).

References

Alfred M. Bruckstein, David L. Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

D. S. Taubman and Michael W. Marcellin. *JPEG 2000: image compression fundamentals, standards and practice.* Kluwer Academic Publishers, 2001.

Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Transactions on Image Processing*, 13(1):1–14, 2004.

Acknowledgements

I would like to thank the Institute for Physical Science and Technology at the University of Maryland, College Park and my advisor John J. Benedetto for the support I received during the preparation of this work.

For further information

Please contact ant@umd.edu (Alfredo Nava-Tudela)

http://www.math.umd.edu/~ant/