Thoughts on Constructing an Image Compression Matrix

W. Ryan Livingston

May 12, 2011

I will try to describe my formulation of the problem of creating a compression matrix to be used as the coefficient matrix in a least squares problem for compressing images. Let's formulate the problem.

Problem 1. Given a data set, $\{b_i\}_{i=1}^m \subset \mathbb{R}^{n^2}$ of vector representation of $n \times n$ blocks of grayscale JPEG images how can one construct a matrix, A, of dimensions $n^2 \times k$ with $k \leq n^2$ so that if $x \in \mathbb{R}^k$ solves the least squares problem

$$\min_x \|Ax - b\|_2$$

then Ax is a good visual approximation to b. Where we are taking b to be an $n \times n$ block of the image being compressed.

We are therefore looking for a matrix A whose range is "close" to most vectors in $\mathbb{R}^{n^2}$ which represent blocks of image data. Here is a particular attempt at a solution to this problem.

Construct the $n^2 \times m$ matrix whose columns are the b_i and call it A'. One interesting feature of this matrix is that it contains every $n \times n$ block of training image data in its range. Therefore, we wish to extract the "important" basis elements of its image in order to construct the matrix A in the above problem.

Recall that the SVD of a matrix provides a way of constructing an orthonormal basis for the image of a matrix. So write $A' = U\Sigma V^T$. Then the left singular vectors corresponding to positive singular values, i.e. the columns of U corresponding to positive singular values, form the desired orthonormal basis of $\operatorname{Im} A'$. Another feature of the SVD is that the singular values assign a weight to each left singular vector. Therefore, we may choose $A = \begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix}$ since $(\Sigma)_{11} \geq (\Sigma)_{22} \geq \cdots \geq (\Sigma)_{\min\{n^2,m\}}$. This has the nice effect of reducing or eliminating the need to worry about redundant data in the b_i. These nearly parallel vectors will give data corresponding to small singular values and therefore will not be used provided a large sample of image data.

Making this choice for A guarantees that it spans the same subspace as the best rank k approximation to A' with respect to the l_2 operator norm. Also, this eliminates the need to compute this rank k approximation thereby saving some numerical error. Finally, by choosing k appropriately, we can adjust the compression ratio.

Another issue which needs to be considered is,

Problem 2. How to choose the training data, $\{b_i\}$?

In my work, I tried to choose images which had a large amount of variety both within a single picture and within the entire set. It seems as though a similar approach will yield more information for less work. Some examples of things I considered were: light foreground/dark background and vice versa, high contrast areas, areas of regular patterns like bricks, areas of irregular objects like landscapes etc.

The question also arises, is more always better or will too much information add to instability? Through some experimentation, some people observed that adding more image data kept increasing perceived image quality so there is a bit of empirical evidence to support this.