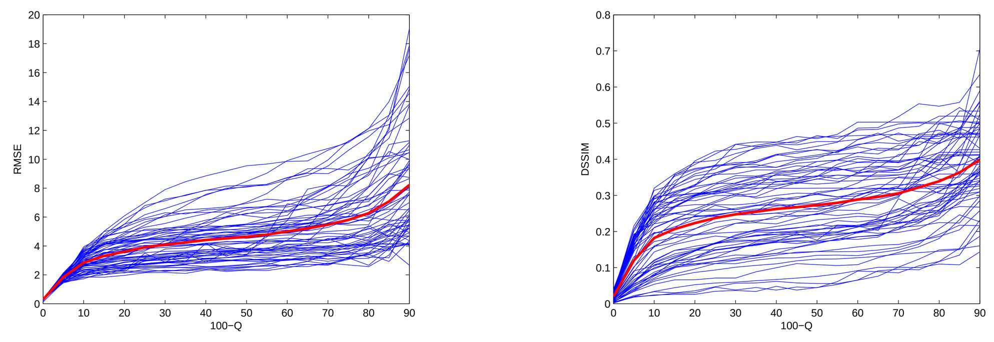
Objective image quality measures of degradation in compressed natural images and their comparison with subjective assessments

Allison K. Cheeseman, Ilona A. Kowalik-Urbaniak and Edward R. Vrscay

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Introduction

This paper is concerned with the degradation in natural images produced by JPEG (and JPEG2000) compression as measured by traditional root mean squared error (RMSE) and the Structural Similarity-based error DSSIM (see below). As the degree of compression increases (i.e., JPEG Quality Factor, Q, decreases), degradation increases:

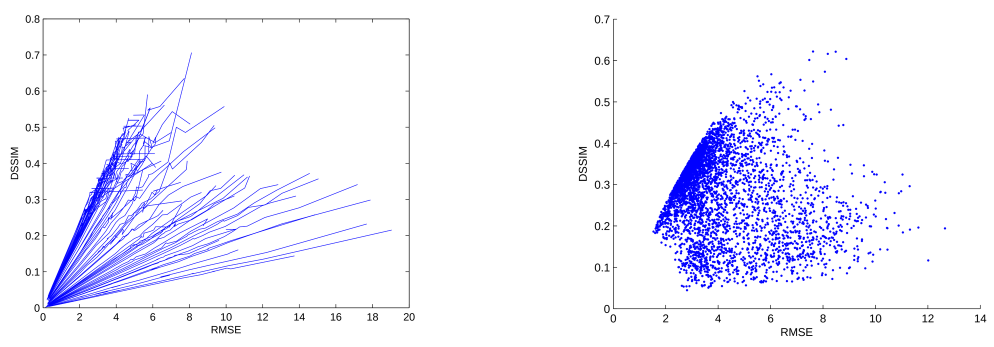


Degradation vs. $Q' = 100 - Q$ for the 256 nonoverlapping 8×8-pixel blocks of *Lena* image. **Left:** RMSE. **Right:** DSSIM. In both cases, the mean values are also plotted (in red).

Our study has been twofold:

- To find relationships between the amount degradation in an image block and its various statistical properties. The goal is to identify blocks that will exhibit lower (**"good"**) or higher (**"bad"**) rates of degradation as the degree of compression increases.
- To compare the above **objective** characterizations with **subjective** assessments of observers.

An additional complication: Blocks that are **bad** w.r.t. one measure, say, RMSE, are not necessarily "bad" with respect to the other, as shown in the plot on the right below.



Left: $RMSE$ vs. $DSSIM$ compression errors for the 256 nonoverlapping 8×8-pixel blocks of the *Lena* image over the range $100 \leq Q \leq 10$. **Right:** $RMSE$ vs. $DSSIM$ errors for all 4096 blocks at $Q = 50$, showing poor correlation between RMSE and DSSIM error measures.

Obvious questions:

1. What, if any characteristics of blocks can be used to separate "bad" blocks from "good" blocks **for a given fidelity measure**? Previously, we examined standard deviation, total variation, low- and high-frequency content **with no significant success**.
2. What, if any, features can be used to characterized blocks that are "bad" with respect to one measure and "good" with respect to the other?
3. Which fidelity measure is "better visually", i.e., which measure corresponds better to human visual perception of degradation?

In this study, we have found that (i) energy and (ii) high-frequency fraction, as opposed to high-frequency content, are better indicators.

Definitions of important quantities

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N \times N}$ denote two $N \times N$-dimensional image blocks. Here, $\mathbf{x}$ will usually represent a block of an uncompressed image and $\mathbf{y}$ the corresponding block of the compressed image. As usual,

$$RMSE(\mathbf{x}, \mathbf{y}) = \sqrt{MSE(\mathbf{x}, \mathbf{y})} = \left[\frac{1}{N^2} \sum_{i,j=1}^{N} (x_{ij} - y_{ij})^2 \right]^{1/2} = \frac{1}{N} \|\mathbf{x} - \mathbf{y}\|_2 \tag{1}$$

and

$$DSSIM(\mathbf{x}, \mathbf{y}) = \sqrt{1 - SSIM(\mathbf{x}, \mathbf{y})}\,, \tag{2}$$

where $SSIM(\mathbf{x}, \mathbf{y})$ denotes the **Structural Similarity Measure** between $\mathbf{x}$ and $\mathbf{y}$ (see paper). Then $DSSIM(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$.

For $N = 8$, let c_{kl}, $0 \leq i, j \leq N - 1$, denote the JPEG DCT coefficients of $\mathbf{x} \in \mathbb{R}^{N \times N}$,

$$\mathbf{c} = \begin{pmatrix} c_{00} & c_{01} & \cdots & c_{07} \\ c_{10} & c_{11} & \cdots & c_{17} \\ \vdots & \vdots & \ddots & \vdots \\ c_{70} & c_{71} & \cdots & c_{77} \end{pmatrix} \qquad \begin{pmatrix} \textbf{low} & \\ \textbf{frequencies} & \\ & \textbf{high} \\ & \textbf{frequencies} \end{pmatrix}. \tag{3}$$

From Parseval's Theorem, $\|\mathbf{c}\|_2 = \|\mathbf{x}\|_2$. Define the counterdiagonal vectors of the $\mathbf{c}$ matrix,

$$\mathbf{d}_m = \{c_{kl},\ k + l = m\}, \quad 0 \leq m \leq 14\,. \tag{4}$$

Furthermore, define the **low-** and **high-frequency content** of block $\mathbf{x}$ to be, respectively,

$$\|\mathbf{x}\|_{lc} = \left[\sum_{m=1}^{6} \|\mathbf{d}_m\|_2^2 \right]^{1/2}, \qquad \|\mathbf{x}\|_{hc} = \left[\sum_{m=7}^{14} \|\mathbf{d}_m\|_2^2 \right]^{1/2}. \tag{5}$$

Note that the DC coefficient $c_{00} = N\bar{\mathbf{x}}$ is omitted from $\|\mathbf{x}\|_{lc}$ since it is generally much greater in magnitude than the other c_{ij}, thereby masking their contributions. Moreover, c_{00} is virtually unchanged by compression. Also note that

$$\|\mathbf{x}\|_{lc}^2 + \|\mathbf{x}\|_{hc}^2 = \|\mathbf{x}\|_2^2 - c_{00}^2 = \|\mathbf{x}_0\|_2^2 = (N^2 - 1)s_{\mathbf{x}}^2 = \textbf{reduced energy of } \mathbf{x}\,. \tag{6}$$

Now define the **low-** and **high-frequency fractions** of an image block $\mathbf{x}$ as, respectively,

$$\|\mathbf{x}\|_{lf} = \frac{\|\mathbf{x}\|_{lc}}{\|\mathbf{x}_0\|_2}, \qquad \|\mathbf{x}\|_{hf} = \frac{\|\mathbf{x}\|_{hc}}{\|\mathbf{x}_0\|_2}. \tag{7}$$

Note that by definition,

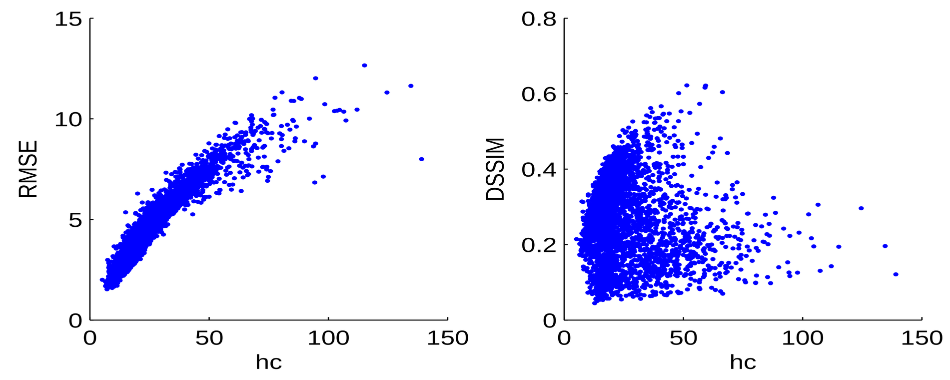
$$\|\mathbf{x}\|_{lf}^2 + \|\mathbf{x}\|_{hf}^2 = 1\,. \tag{8}$$

Eq. (7) provides a more block-independent, hence compact, characterization of low- and high-frequency content than Eq. (5).

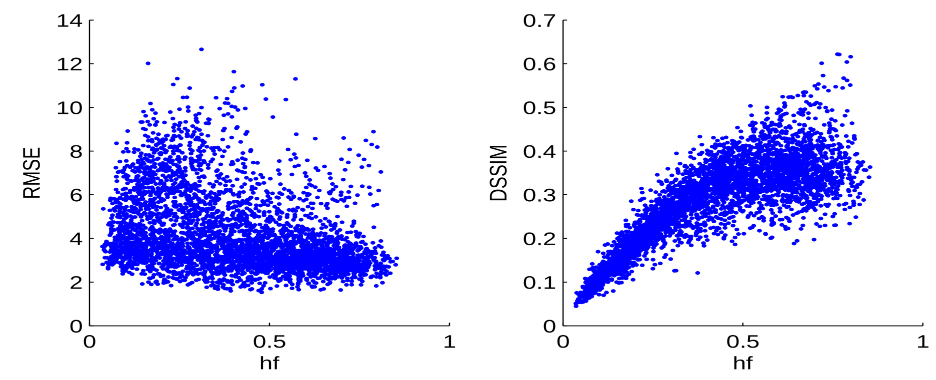
Quantitative measure of compression-induced degradation of image blocks

With regard to the RMSE and DSSIM distances between uncompressed and compressed (8×8-pixel) blocks, we present the following evidence in the paper:

1. For a given Q, blocks with low reduced energy (E) (as well as total variation (TV) and high-frequency content (HC)) exhibit lowest degradation in terms of RMSE. This is expected from a knowledge of the action of JPEG on higher-frequency DCT coefficients which generally have lower magnitudes. However, there is a larger spread of DSSIM errors in the low TV, E and HC regimes as shown in the plots below. This indicates that **not all blocks with low RMSE are necessarily visually "good" in the sense of DSSIM**.

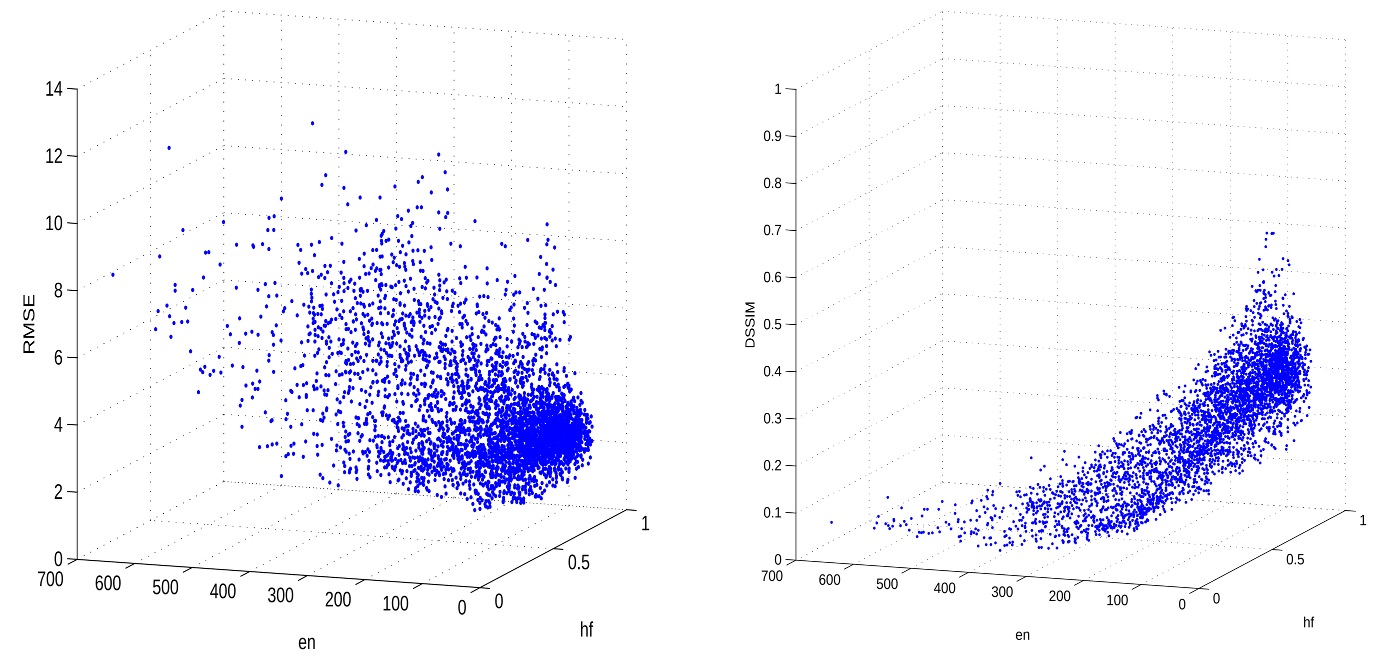


2. **High-frequency fraction**, as opposed to **high-frequency content**, is found to be strongly correlated to DSSIM error, as shown below,



This can once again be understood in terms of the action of JPEG compression in removing higher-frequency DCT coefficients.

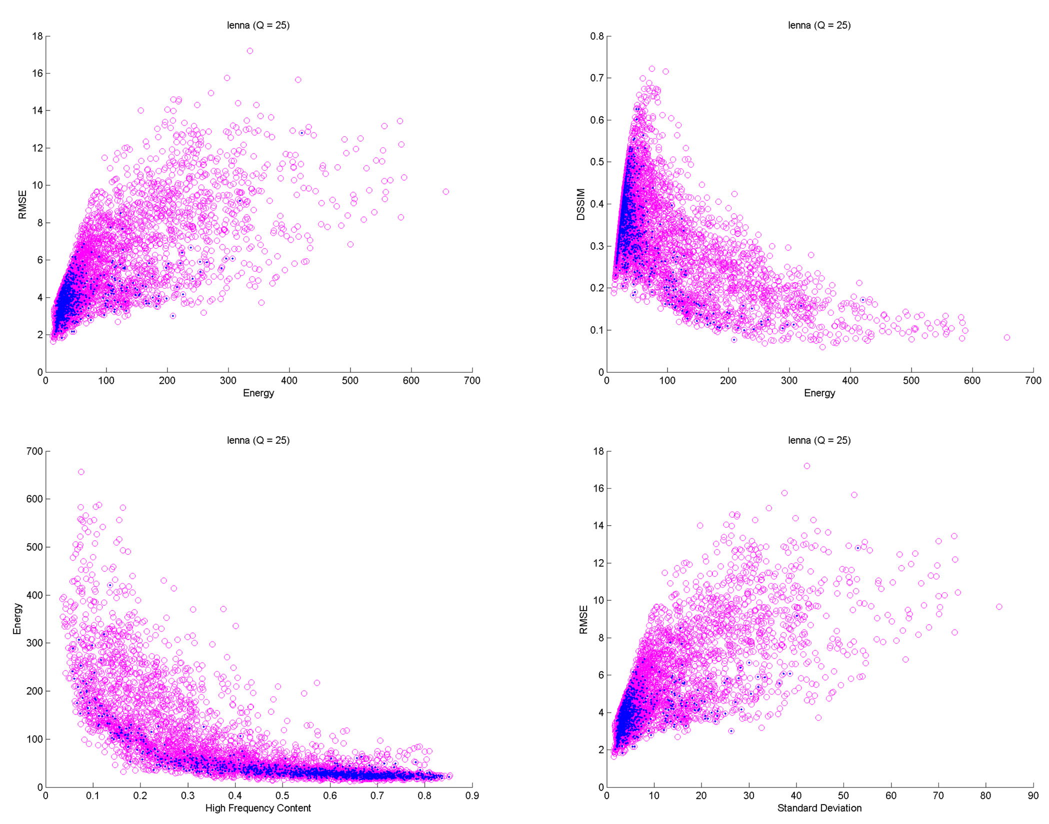
In order to achieve even more separability of low RMSE/high DSSIM blocks, we plot RMSE and DSSIM errors vs. high-frequency fractions and reduced errors of blocks below.



RMSE and DSSIM distances between 4096 compressed and uncompressed 8×8-pixel blocks of *Lena* image vs. high-frequency fractions (hf) and reduced energies (en) of the blocks.

However, the question remains: Is DSSIM error a good indicator of **visual degradation?**

Comparison with subjective evaluations



Results of subjective analysis of JPEG-compressed *Lena* image at $Q = 25$. Red circles denote all 4096 8×8-pixel blocks of image. Blue dots indicate blocks identified by subject as degraded.

Most noteworthy observations:

1. The "bad" blocks identified by subjects as visually degraded had **low RMSE** errors and **medium-to-high DSSIM** errors.
2. These "bad" blocks correspond to uncompressed blocks with **low reduced energies** and **medium-to-high frequency fractions**.

Concluding comments

1. **The fact that blocks with low RMSE compression error were identified as "visually bad" clearly implies that RMSE is not a good indicator of visual quality/degradation.**
2. **The fact that "visually bad" blocks correspond to uncompressed blocks with low reduced energy/standard deviation indicates a kind of perceptual Weber law for compression: Distortions are more likely to be observed for blocks of lower variance. This is an example of "contrast masking."**
3. **Similar behaviour is observed for JPEG2000 compression and will be reported elsewhere.**