THEORETICAL CAPACITY MEASURES FOR DATA HIDING IN COMPRESSED IMAGES *

Mahalingam Ramkumar and Ali N. Akansu
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
New Jersey Center for Multimedia Research
University Heights, Newark, NJ 07102.
ali@megahertz.njit.edu

ABSTRACT

We present an information-theoretic approach to obtain an estimate of the number of bits that can be hidden in still images, or, the *capacity of the data-hiding channel*. We show how the addition of the message signal or signature in a suitable transform domain rather than the spatial domain can significantly increase the channel capacity. Most of the state-of-the-art schemes developed thus far for data-hiding have embedded bits in some transform domain, as it has always been implicitly understood that a decomposition would help. In this paper we compare the achievable data-hiding capacities for different decompositions like DCT, Hartley, Hadamard, and subband transforms. We show that transforms with inferior energy compaction property like Hartley and Hadamard are better choices for the decomposition, than transforms with good energy-compaction property, like DCT or subband (wavelet) transforms.

Keywords: I nformation Hiding, Channel Capacity, Watermarking, Still Images.

1. INTRODUCTION

The fast growth of digital networks, and the ever-decreasing cost of computers, printers and digital transmission have made digital media increasingly popular over the conventional analog media. However, digital media also causes extensive opportunities for mass piracy of copyrighted material. It is therefore very important to have ways and means to detect copyright violations and control access to digital media.

Data-hiding or steganography, is a rapidly growing field with potential applications for copyright protection (watermarking), hiding executables for access control of digital multimedia data, embedded captioning, secret communications, and others. It is therefore of significant interest to have a theoretical estimate of the number of bits that can be hidden in multimedia data. We provide an information-theoretic approach to estimate the number of bits that can be hidden in still images.

Data-hiding schemes (in still images) can be broadly classified into two categories. The first category is called *cover image escrow* hiding techniques, where, the original image is needed for extracting the hidden information.[1] The second category is the *oblivious detection* techniques,[2] where the original image is *not* required for extraction of the embedded message or signature. However, the schemes in the first category are of limited use. For instance, the cover image escrow hiding schemes may not resolve rightful ownership.[2,3] In addition, the receiver does not have access to the original image most of the time.

Early work in data-hiding mainly consisted of modifying the least significant bits (LSB) of data to embed some message bits. However, it is usually necessary for the embedded bits to survive common signal processing operations. Unfortunately these methods lack the necessary robustness, to make them useful for many applications. Most of the current techniques for data-hiding in images utilize some decomposition for embedding the message bits. Among different orthonormal decomposition techniques, it was probably the inspiration from image compression applications that caused DCT and subband (wavelet) transforms to be more popular than the others. In this paper, we show why a decomposition helps to improve the data-hiding capacity. We also compare the achievable capacities for different decompositions like Hartley, DCT, Hadamard and subband (wavelet) transforms.

*This work was partially supported by Panasonic Technologies Inc., Plainsboro, NJ.

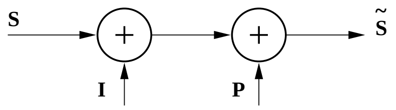

Figure 1. The Data Hiding Channel.

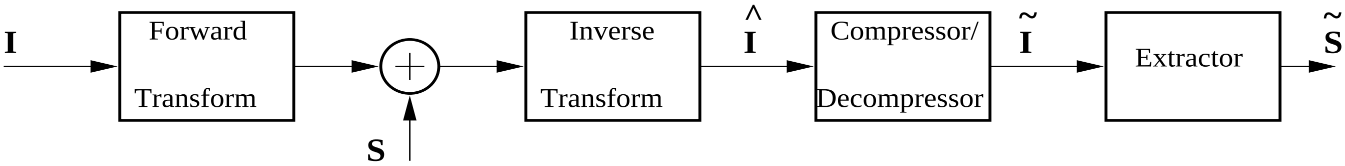

Figure 2. Generalized Schematic of Data Hiding / Retrieval.

2. PROBLEM STATEMENT

Let $\boldsymbol{I}$ be the original (cover) image, to which a message $\boldsymbol{S}$ (a representation for embedded information bits) is added, such that

$$\hat{\boldsymbol{I}} = \boldsymbol{I} + \boldsymbol{S}. \tag{2.1}$$

The modified image $\hat{\boldsymbol{I}}$, is *visually indistinguishable* from $\boldsymbol{I}$ and may typically be subjected to a lossy compression scheme, like JPEG

$$\tilde{\boldsymbol{I}} = \mathcal{C}(\hat{\boldsymbol{I}}), \tag{2.2}$$

where $\mathcal{C}(.)$ denotes the compression / decompression operation. Therefore, embedded bits in image $\boldsymbol{I}$ are to be extracted from $\tilde{\boldsymbol{I}}$. We would like to know the maximum number of bits that can be hidden and recovered from the image with an arbitrarily low probability of error, namely, the *capacity of the data-hiding channel*, for a given compression scenario.

A block diagram of the data-hiding channel is shown in Figure 1, where $\boldsymbol{S}$ is the message (signature) to be transmitted through the channel. Note that there are two sources of noise; $\boldsymbol{I}$, the noise due to the (original) cover image, and $\boldsymbol{P}$, the noise component due to processing (compression / decompression). Hence, $\tilde{\boldsymbol{S}}$ is the "corrupted" message. Note that for the cover image escrow schemes, there is only one source of noise - due to processing. The image noise can be subtracted from the received image $\tilde{\boldsymbol{I}}$. One can expect such schemes to have higher capacity than the oblivious detection schemes.

Figure 2 displays the block diagram of a typical data-hiding scheme. In this paper, we assume the system of Figure 2 to calculate the capacity of data-hiding channel. An information theoretic approach for the capacity measure of data-hiding channels was reported in Smith *et. al.*[4] The study however, was limited in scope, since it was assumed that the message is added to the original image in the spatial domain. We show in this paper how the capacity of the data-hiding channel can be improved by using a suitable transform.

3. CAPACITY OF ADDITIVE NOISE CHANNELS

Prior to considering the data-hiding channel of Figure 1, we consider the simpler channel displayed in Figure 3(a). $\boldsymbol{X} \sim [f_X(x), \sigma_x^2]$ is the message to be transmitted, $\boldsymbol{Z} \sim [f_Z(z), \sigma_z^2]$ is the additive noise in the channel, and $\boldsymbol{Y} \sim [f_Y(y), \sigma_y^2]$ is the received signal at the output of the channel, along with their pdfs and corresponding variances. We also assume that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent, implying that $\sigma_y^2 = \sigma_x^2 + \sigma_z^2$. Therefore, the channel capacity is given by[5]

$$\boldsymbol{C} = \max_{f_X(x)} I(\boldsymbol{X}, \boldsymbol{Y}) = \max_{f_X(x)} h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}) = \max_{f_X(x)} h(\boldsymbol{Y}) - h(\boldsymbol{Z})\,\text{bits}. \tag{3.3}$$

where $I(\boldsymbol{X}, \boldsymbol{Y})$, is the *mutual information* between $\boldsymbol{X}$ and $\boldsymbol{Y}$. For a given noise statistics $f_Z(z)$ and input variance σ_x^2, one can maximize the entropy of the output $\boldsymbol{Y}$,

$$h(\boldsymbol{Y}) = -\int f_Y(y)\log_2(f_Y(y))\mathrm{d}y \text{ bits}, \tag{3.4}$$

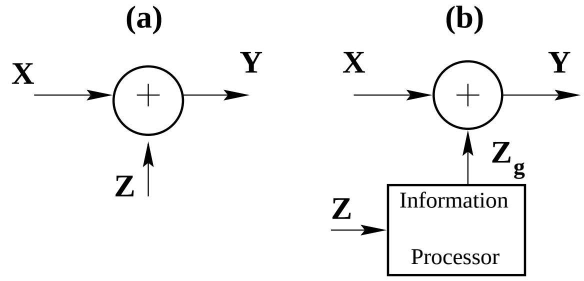

Figure 3. (a) A Simple Additive Noise Channel. (b) The Channel of (a) Modified to Obtain Equivalent Additive Gaussian Noise.

by choosing a suitable distribution $f_X(x)$ for the input message $\boldsymbol{X}$. For a given variance σ_y^2, the maximum entropy value of $h(\boldsymbol{Y}) = \frac{1}{2}\log_2(2\pi e\sigma_y^2)$ bits is achieved when $\boldsymbol{Y}$ has a normal distribution. For instance, the maximum entropy value is achievable if both pdfs $f_Z(z)$ and $f_X(x)$ are normally distributed. However, for an arbitrary distribution $f_Z(z)$, and a fixed σ_x^2, the maximum achievable entropy value is not immediately obvious.

To calculate that, we pass the noise $\boldsymbol{Z}$ through an ideal *information processor*, (see Figure 3(b)) which does not alter the amount of information in $\boldsymbol{Z}$, but changes its statistics to a Gaussian distribution for its output $\boldsymbol{Z}_g$. Since the output of the information processor has the same entropy as the input, the variance of the output, σ_{zg}^2, can be obtained by solving the equation

$$h(\boldsymbol{Z}_g) = h(\boldsymbol{Z}) = \frac{1}{2}\log_2(2\pi e\sigma_{\text{zg}}^2)\text{ bits.} \tag{3.5}$$

It is known that the Gaussian distribution has the highest entropy for a given variance.[5] Alternately, the Gaussian distribution has the least variance for a given entropy. Thus it is always true that $\sigma_{\text{zg}}^2 \le \sigma_z^2$. We call σ_{zg}^2 the *entropy equivalent Gaussian variance*. The maximum value of $h(\boldsymbol{Y})$ is therefore obtained as

$$\max_{f_X(x)} h(\boldsymbol{Y}) = \max_{f_X(x)} h(\boldsymbol{X} + \boldsymbol{Z}_g) = \frac{1}{2}\log_2(2\pi e(\sigma_{\text{zg}}^2 + \sigma_x^2))\text{ bits.} \tag{3.6}$$

In order to calculate the channel capacity, we can now replace $f_Z(z)$ by $N[0, \sigma_{\text{zg}}^2]$.

$$\boldsymbol{C} = \max_{f_X(x)} h(\boldsymbol{Y}) - h(\boldsymbol{Z}_g) = \frac{1}{2}\log_2(1 + \frac{\sigma_x^2}{\sigma_{\text{zg}}^2})\text{ bits.} \tag{3.7}$$

Note that the two channel noise sources given in Figure 1 can be replaced by a single Gaussian noise source with the combined variance of $\sigma_{\text{ig}}^2 + \sigma_p^2$, where σ_{ig}^2 is the equivalent Gaussian variance for the image noise $\boldsymbol{I}$, and σ_p^2 is the variance of the processing noise. If σ_s^2 is the message signal energy, the capacity of the data-hiding channel can be expressed as

$$\boldsymbol{C}_h = \frac{1}{2}\log_2(1 + \frac{\sigma_s^2}{\sigma_{\text{ig}}^2 + \sigma_p^2})\text{ bits.} \tag{3.8}$$

As the first approach to calculate the capacity of the data-hiding channel, we assume identity transformation for the Forward and Inverse Transform blocks in Figure 2. The image noise $\boldsymbol{I}$ is due to the original image pixels, which are assumed to be uniformly distributed random variables u taking values between 0 and 255 with variance σ_i^2. Let σ_p^2 be the variance of the noise (per pixel) introduced due to processing, (*e.g.* compression). Since the processing noise is usually a result of many independent operations, we shall call upon the Central Limit Theorem,[6] and assume a Gaussian distribution for the processing noise. Finally, let σ_s^2 be the average energy per pixel allowed for the message. If MN is the number of pixels in an image, then the energy (or variance) of the zero-mean message signal is calculated as

$$\sigma_s^2 = \frac{\sum_{i=1}^{MN} S_i^2}{MN}, \tag{3.9}$$

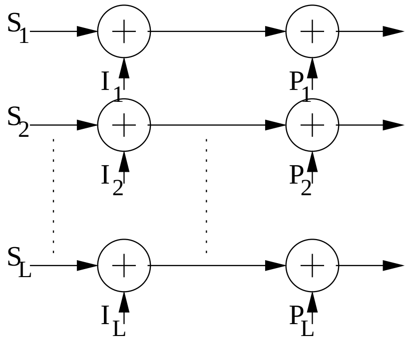

Figure 4. Decomposition of the Data-Hiding Channel into Parallel Sub-Channels

where, S_i is the message signal added to the i^{th} pixel. The (differential) entropies, $h(g)$, of a Gaussian random variable g, with variance of σ_g^2, and $h(u)$, that of a uniformly distributed random variable u with variance σ_u^2 are expressed as[5]

$$\begin{aligned} h(g) &= \frac{1}{2}\log_2(2\pi \mathrm{e}\sigma_g^2)\,\text{bits} \\ h(u) &= \frac{1}{2}\log_2(12\sigma_u^2)\,\text{bits}. \end{aligned} \tag{3.10}$$

From Eq. 3.10, the *entropy equivalent Gaussian noise* (or the Gaussian random variable that has the same entropy as the uniform random variable u of variance σ_i^2), has a variance given by

$$\sigma_{\text{ig}}^2 = \frac{12}{2\pi \mathrm{e}}\sigma_i^2. \tag{3.11}$$

In order to be more explicit, let us derive the capacity of the data hiding channel quantitatively. We would expect the variance of u, the pixel values, to be given by $\sigma_i^2 = \frac{255^2}{12}$ (or $\sigma_i = 73.6$). However, statistics from many test images show that $\sigma_i = 55$. Therefore, we assume that u has a uniform distribution with $\sigma_i = 55$. From Eq. (3.11) it is calculated that $\sigma_{\text{ig}} = 55(\frac{12}{2\pi e})^{0.5} \approx 46$. If we allow a degradation of the image after the addition of a message to a PSNR of 40 dB, then the message energy is calculated to be $\sigma_s^2 = 6.5$. Furthermore, if the image goes through JPEG compression at 50% quality, then it is measured for test images that the processing noise has a standard deviation of $\sigma_p \approx 6.7$. This would yield a capacity $\boldsymbol{C}_h$ value of 0.0022 bits/pixel (140 bits for a 256×256 image). Even if the message-embedded image undergoes some other processing which reduces the PSNR to 22 dB (the image would be barely recognizable), where $\sigma_p \approx 20$, the capacity $\boldsymbol{C}_h$ would still be 0.0019 bits per pixel (about 124 bits for a 256×256 image).

One can see that hiding the message in the image domain can be very robust. However, in most cases, we do not require such robustness. Since most data-hiding applications aim to protect and ascertain copyright or control access, it is unlikely in such a scenario that anyone would want to claim ownership or control access of an image of no commercial value (an image which has been significantly degraded in perceptual quality). Typically, it is sufficient if the message survives well-known image compression/ decompression operations with acceptable quality.

Given that we are satisfied with less robustness than the above mentioned scheme offers, could do better than this? It is intuitive that a decomposition of the image into its different frequency bands might help. In Figure 4, the channel of Figure 1 is decomposed into its multiple sub-channels.

The decomposition is performed by the Forward and Inverse Transform blocks of Figure 2. The decomposition of an image into its L sub-bands results in L parallel sub-channels with two noise sources in each sub-channel. Let $\sigma_{i_j}^2$, $j = 1 \cdots L$, be the variances of the coefficients for each sub-band (or the variances of the image noise in each sub-channel) of the decomposition. Similarly, let their corresponding equivalent Gaussian variances be $\sigma_{\text{ig}_j}^2$. If $\sigma_{p_j}^2$ is the variance of the processing noise (Gaussian) in the j^{th} sub-channel, then, the combined total capacity of the L

parallel sub-channels is given by

$$\boldsymbol{C}_h = \frac{MN}{2L}\sum_{j=1}^{L}\log_2(1+\frac{v_j^2}{\sigma_{\mathrm{ig}_j}^2+\sigma_{p_j}^2})\,\mathrm{bits} \tag{3.12}$$

for an image of size MN pixels. In Eq. (3.12), v_j is the *visual threshold* of band j. In other words, v_j^2 is the maximum message signal energy permitted in band j based on its perceptual quality effects.

We expect the low frequency bands of the decomposition to very noisy due to the high energy content of the image. On the other hand, high frequency components would be very vulnerable to processing, as most compressors would discard them at low bit-rates. At mid-frequency bands, however, we could strike a compromise. If such a decomposition helps, then what is the ideal decomposition? In the following sections, we evaluate the capacity of the data-hiding channel for DCT, Hartley, Hadamard, and uniform subband decomposition based embedding schemes.

4. MODELING CHANNEL NOISE

In order to model the channel noise (the two noise sources $\boldsymbol{I}$ and $\boldsymbol{P}$ in Figure 1), we measure their statistics from 10 monochrome test images of size 256×256, and their JPEG and SPIHT[7] compressed versions at given bit rates. The cover images are decomposed into L sub-bands using an orthonormal transform. Let $f_{I_j}(i_j)$ be the distribution of the j^{th} sub-band with variance $\sigma_{i_j}^2$. (The image noise $\boldsymbol{I}$ is split into its components in L sub-channels, which are modeled as random variables $f_{I_j}(i_j)$ with variances $\sigma_{i_j}^2$, $j = 1\cdots L$.)

Having obtained the variances of the image noise in each sub-channel, the next step is to obtain their equivalent Gaussian variances. This is achieved by plotting a histogram of the coefficients for each band, and calculating the entropy. If Δx is the width of the n bins of the histogram $g_j(m)$, $m = 1\cdots n$, and p is the total number of coefficients in band j, the entropy $\mathcal{H}_j$ and the equivalent Gaussian variance $\sigma_{\mathrm{ig}_j}^2$ of the sub-band are obtained as

$$\begin{aligned}\mathcal{H}_j &= -\sum_{i=1}^{n}\frac{g_j(i)}{p\Delta x}\log_2(\frac{g(i)}{p\Delta x})\Delta x,\ \mathrm{bits}\\ \sigma_{\mathrm{ig}_j}^2 &= \frac{2^{2\mathcal{H}_j}}{2\pi\mathrm{e}}.\end{aligned} \tag{4.13}$$

Thus, the image noise in sub-channel (band) j can be substituted by a Gaussian noise of variance $\sigma_{\mathrm{ig}_j}^2$.

Let the compression noise in each sub-channel be $\sigma_{p_j}^2$, $j = 1\cdots L$. As in the previous section, it is justified to assume a Gaussian distribution for the processing noise for each channel. We obtain $\frac{MNn_i}{L}$ samples for each sub-band from n_i test images. Let i_{j_k}, $k = 1,\ldots,\frac{MNn_i}{L}$, be the coefficients of band j. Let $\tilde{i}_{j_k}$, $k = 1,\ldots,\frac{MNn_i}{L}$ be the corresponding coefficients for the images subjected to some lossy compression scheme.

It can be easily seen that the processing noise in each sub-band *can not* be obtained as $\tilde{i}_{j_k} - i_{j_k}$. Consider a scenario, where DCT is used for the decomposition, and low quality JPEG for processing. Let us assume that a high frequency sub-band is completely removed due to compression ($\tilde{i}_{j_k} = 0\,\forall\, k$ for some j). This implies that all information buried in that sub-channel (sub-band) is lost. In other words, the processing noise in that sub-channel has *infinite* variance. This is because no *correlation* exists between $\tilde{i}_{j_k}$ and i_{j_k}. We therefore obtain the equivalent additive noise in each sub-channel as a noise uncorrelated with $\boldsymbol{i}_j$, that would cause the same *reduction in correlation* between $\boldsymbol{i}_j$ and $\tilde{\boldsymbol{i}}_j$. We define the intra-band correlation as

$$\frac{\langle \boldsymbol{i}_j, \tilde{\boldsymbol{i}}_j\rangle}{|\boldsymbol{i}_j||\tilde{\boldsymbol{i}}_j|} = \frac{\langle \boldsymbol{i}_j, (\boldsymbol{i}_j+\boldsymbol{n}_j)\rangle}{|\boldsymbol{i}_j||\boldsymbol{i}_j+\boldsymbol{n}_j|} = \rho_j, \tag{4.14}$$

where $\boldsymbol{n}_j$ is a vector of (zero mean) Gaussian random variables which is uncorrelated with $\boldsymbol{i}_j$. Then, $\sigma_{n_j}^2 = |\boldsymbol{n}_j|^2$ is the variance of the *equivalent additive noise due to compression* (or $\sigma_{p_j} = \sigma_{n_j}$). Since $\langle \boldsymbol{i}_j, \boldsymbol{n}_j\rangle = 0$, Eq. (4.14) can be simplified to obtain

$$\sigma_{p_j}^2 = |\boldsymbol{n}_j|^2 = (\frac{1}{\rho_j^2}-1)|\boldsymbol{i}_j|^2 \tag{4.15}$$

Note that in Eq. (4.15) when $\rho_j \to 0$, $\sigma_{p_j} \to \infty$.

5. VISUAL THRESHOLD

The value of the *visual threshold* for sub-channel j, v_j in Eq. (3.12) however, is highly subjective. Since the amount of message signal energy permitted in any sub-band is determined by the visual threshold, different models for visual thresholds would yield different estimates of achievable capacity. Since it is well known that the human visual system is more sensitive to the lower frequencies than the higher frequencies, the signal-to-noise-ratio (message signal to image noise) should be smaller for lower frequency sub-bands. In general lower frequency sub-bands have higher variances. Hence, we choose the visual threshold v_j as

$$v_j^2 = K(\sigma {i_j}^2)^\alpha \tag{5.16}$$

where $0 < \alpha < 1$, and $K << \sigma_{i_j} \, \forall j$, is a constant. When $\alpha = 0$, the message signal energy is distributed equally among all sub-bands regardless of their variances. On the other hand, when $\alpha = 1$ the message signal energy is distributed in the ratio of the band variances.

From Eqs. (3.12) and (5.16), for the case of *no processing noise*, if we assume that all sub-channels have the same pdf type (such that $K\sigma i_j = K_1 \sigma_{\mathrm{ig}_j}$), the channel capacity can be calculated as

$$\boldsymbol{C}_h = \frac{MN}{2L} \sum_{j=1}^{L} \log_2(1 + \frac{K_1 \sigma_{\mathrm{ig}_j}^{2\alpha}}{\sigma_{\mathrm{ig}_j}^2}) \approx \frac{MN}{2L} \log_2\left(1 + \sum_{j=1}^{L} \frac{K_1}{\sigma_{\mathrm{ig}_j}^{2(1-\alpha)}}\right), \tag{5.17}$$

Note that for the case of $\alpha = 1$, the decomposition does not have any effect on the capacity. However, for $\alpha < 1$, $\boldsymbol{C}_h$ can be increased by choosing a suitable transform, as shown in the next section. Thus, the increase in capacity is due to the fact that one can add *relatively* more message signal energy to bands of lower variances (or high frequency bands).

6. CHANNEL CAPACITY VS CHOICE OF TRANSFORM

It should be noted that both Eqs. (3.12) and (5.17), are subject to the following constraints

$$\begin{aligned} \sum_{j=1}^{L} \sigma_{i_j}^2 &= L\sigma_i^2 \\ \sum_{j=1}^{L} \sigma_{\mathrm{ig}_j}^2 &= L\sigma_{\mathrm{ig}}^2 \\ \mathcal{I} &= \frac{1}{2}\log_2(2\pi e \sigma_{\mathrm{ig}}^2) \end{aligned} \tag{6.18}$$

where σ_i^2 is the variance of images, σ_{ig}^2 is the entropy equivalent Gaussian variance for σ_i^2, and $\mathcal{I}$ is the average entropy of images. In other words, the above constraints state that an unitary transform preserves the total energy and the total information content, or the entropy. With the above constraints, it can be shown that the *minimum* channel capacity (for the case of *no processing noise* or Eq.(5.17)) is achieved for $\sigma_{\mathrm{ig}_j} = \sigma \, \forall j$, or when no decomposition (spatial embedding) is used.

Note that a transform with good energy compaction or higher Transform Coding Gain (GTC)[8] would result in more *imbalance* of the coefficient variances. This would enhance the term $\sum_{j=1}^{L} \frac{K_1}{\sigma_{\mathrm{ig}_j}^{2(1-\alpha)}}$ in Eq. (5.17), and therefore increase the capacity (when the processing noise is small). Therefore, good energy compaction transforms like DCT and subband transforms are good embedding decompositions for low processing noise scenarios. However, we should expect that the *reduction in capacity* with *increase in processing noise* to be lower for transforms like Hadamard and Hartley, which are unsuitable for compression purposes. While JPEG compression at low quality is certain to remove almost all the high frequency components of DCT coefficients, it will not affect the high frequency DFT and Hadamard coefficients to the same extent. Thus decompositions unsuitable for compression would in general be more immune to processing noise than decompositions with high GTC, for data hiding purposes.

The distribution of the image and processing noise among various channels (bands) of DCT and Hartley transforms are shown in Figure 5 for JPEG (at 30% quality). It is readily seen that the low image noise (or high frequency)

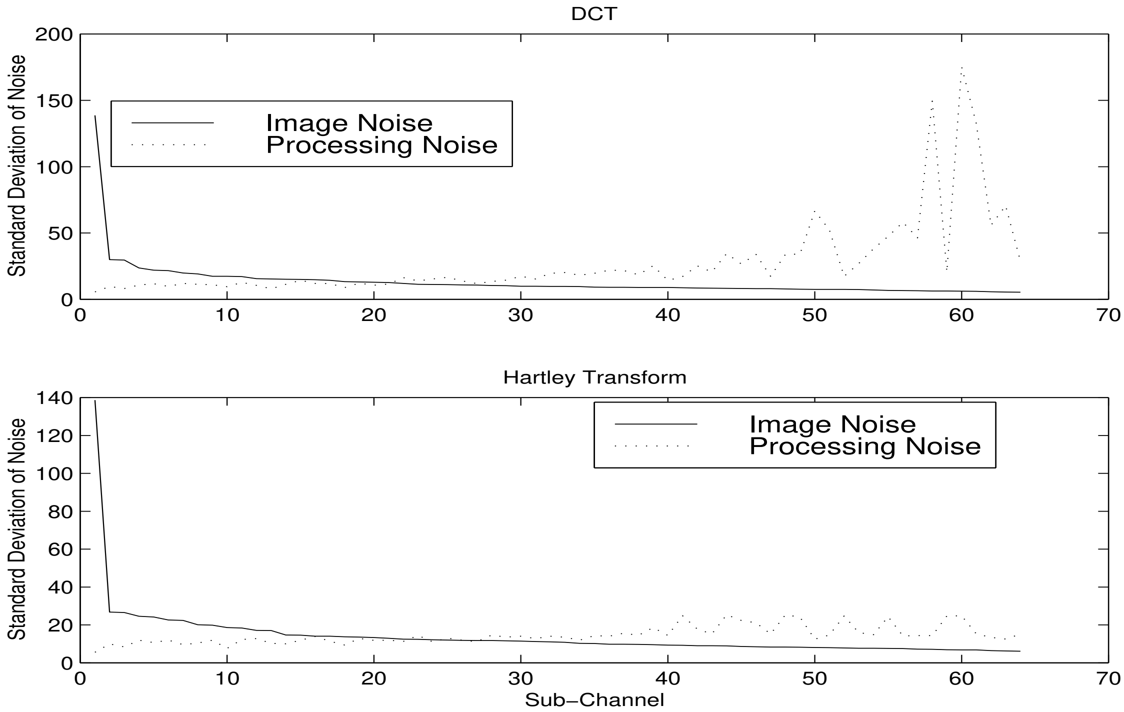

Figure 5. Standard Deviation Image and Processing Noise for the 64 Channels of DCT and Hartley Decompositions.

DCT bands suffer from very high processing noise. This results in very few bands (or channels) in which both image and processing noises are moderate. On the other hand, the high frequency (low image noise) Hartley bands are not affected as badly as the high frequency DCT bands. This results in many 'useful' Hartley transform channels.

The next question that arises is the choice of the number of bands for the decomposition. From Eq. (5.17) we see that a decomposition will not hurt. At worst, it may cause no improvement. Therefore decomposing each sub-channel of say a 16 band decomposition further into four sub-channels can only improve the capacity of data hiding, in this theoretical context.

7. EXPERIMENTS AND CAPACITY BOUNDS

We calculated the coefficient statistics σ_{i_j} for various decompositions like 4×4, 8×8, 16×16 and 32×32 size DCT, Hartley, Hadamard and 16, 64, 256 and 1024 band uniform subband (wavelet, using 8-tap Daubechies filter) decompositions, and σ_{p_j} for JPEG (quality factors 20-75), and SPIHT image compression schemes at rates 1, 0.5, and 0.25 bpp. The set of 10 monochrome test images with 256×256 pixels included Lena, Baboon, Barbara, Goldhill, Airplane, Peppers and Boats.

The theoretical channel capacities for different decompositions (for 256×256 images, or 65536 pixels) like DCT, Hartley (Har), subband (SB), and Hadamard (Had) transformations, are displayed in Figures 6 and 7 for 64 and 256 sub-bands respectively. We use the visual threshold model of Eq.(5.16) with $\alpha = 0.5$ in these figures. The capacities are shown for four different processing noise scenarios, namely,

(a) no processing noise,

(b) processing noise statistics measured for the SPIHT 1 bpp and JPEG-50 cases. This implies that we choose the worst processing noise in each sub-band calculated for SPIHT 1 bpp and JPEG-50. (Choosing the worst ensures that the hidden message would survive SPIHT 1 bpp or JPEG-50.)

(c) processing noise for SPIHT 0.5 bpp and JPEG-30, and

(d) processing noise for SPIHT 0.25 bpp and JPEG-20.

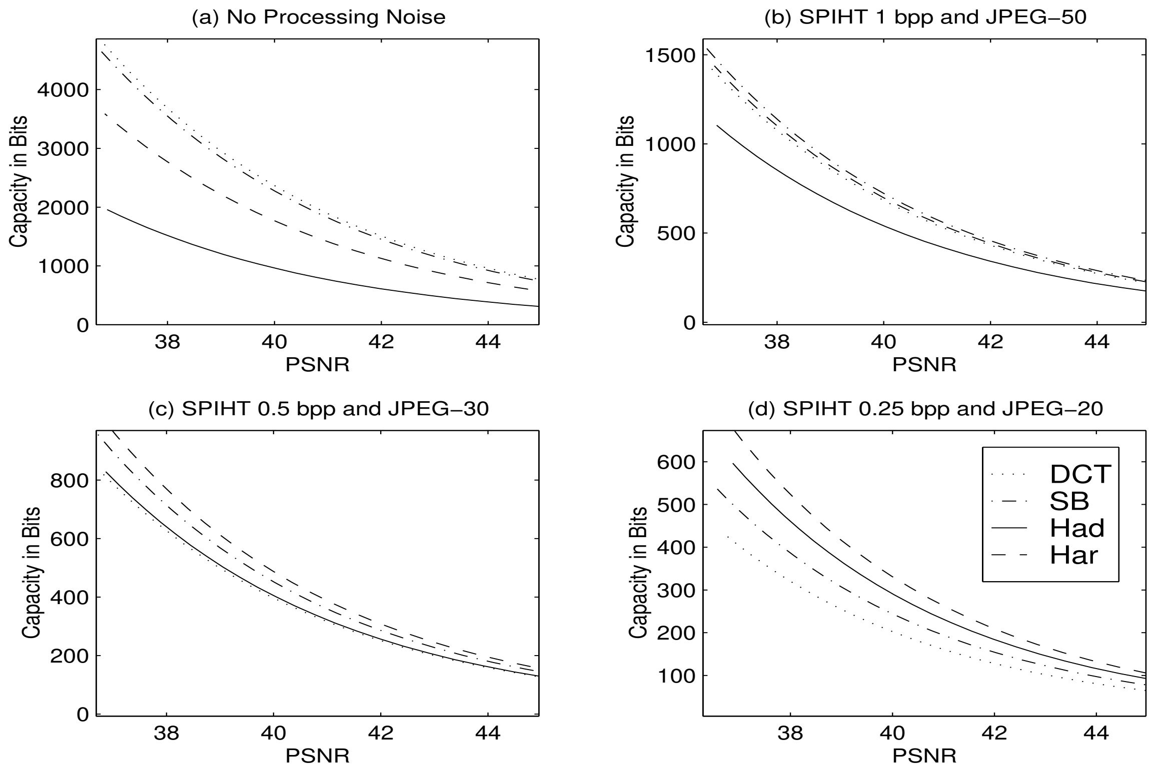

Figure 6. Channel Capacity for 64-band decompositions. (a) No processing Noise, (b) Processing noise from SPIHT 1 bpp and JPEG-50, (c) Processing Noise from SPIHT 0.5 bpp and JPEG-30 and (d) Processing Noise from SPIHT 0.25 bpp and JPEG-20.

From the plots in Figures 6 and 7, we can see that the bit-rates for all decompositions fall with increased processing noise, as expected. Similarly, we see that DCT and subband decompositions are better than Hartley and Hadamard decompositions for detection of the message *when there is no processing noise.* It is also seen that decompositions unfavorable for compression (Hartley and Hadamard) are more immune to processing noise than the decompositions suitable for compression (DCT, subband).

As expected, we also see that the channel capacities increase with an increase in the number of bands of the decomposition. However, the increase in capacity is marginal when processing noise is high.

We can define a *figure of merit*, for each of the L ($\frac{L}{2}+2$ for DFT II) sub-channels for the various decompositions. The figure of merit is given as the ratio of the capacity of each sub-channel to the logarithm of the power of the message signal in that sub-channel. The approximate (rounded) values of the figure of merit for the channels of different decompositions (when the message has to survive JPEG-25 or SPIHT compression at 0.5 bpp), are listed in Table 1 for various 64-band decompositions. These figures indicate the relative performance of each sub-channel, and would therefore be useful in designing hidden communication schemes to make optimal trade-offs between the visual quality of the image and the number of bits that can be embedded. As the figure of merit is normalized with respect to the message signal energy in each band, it is independent of the model used for the visual threshold. The higher figures of merit for the channels of Hartley decomposition shows that it would perform better than other three decompositions for any message signal energy assignment scheme (model for visual threshold). As explained in the previous section, one can immediately see that there exists fewer low efficiency channels for the Hartley decomposition when compared to DCT or subband decompositions.

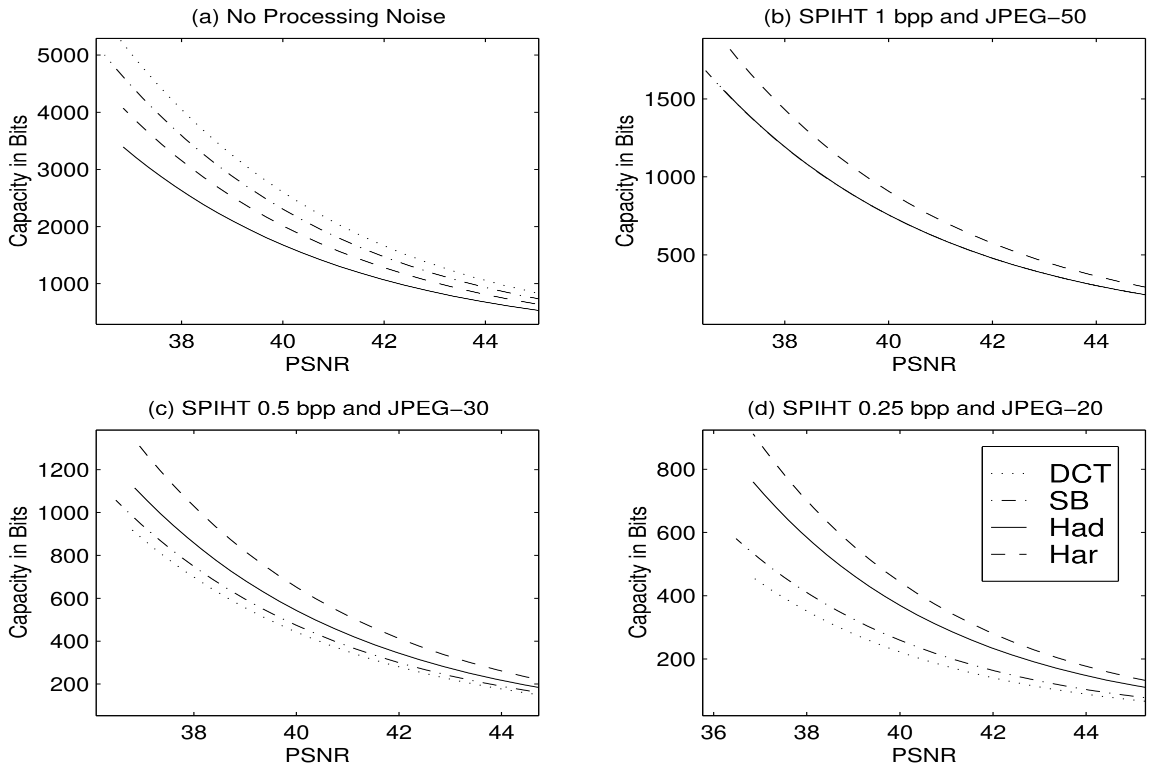

Figure 7. Channel Capacity for 256-band decompositions. (a) No processing Noise, (b) Processing noise from SPIHT 1 bpp and JPEG-50, (c) Processing Noise from SPIHT 0.5 bpp and JPEG-30 and (d) Processing Noise from SPIHT 0.25 bpp and JPEG-20.

(a)-Hartley

0	15	34	37	35	26	17	9
15	23	36	34	40	30	38	26
34	36	31	12	13	14	33	37
37	34	12	2	9	12	25	39
35	40	13	9	39	24	34	48
26	30	14	12	24	13	24	35
17	38	33	25	34	24	34	38
9	26	37	39	48	35	38	25

(b)-DCT

0	8	19	29	37	42	29	23
8	17	28	34	41	28	10	28
19	28	36	40	35	15	7	22
29	34	40	40	23	8	2	22
37	41	35	23	15	2	11	2
42	28	15	8	2	0	0	0
29	10	7	2	11	0	0	3
23	28	22	22	2	0	3	4

(c)-Subband

0	9	29	37	43	41	37	33
9	18	19	26	37	43	32	18
29	19	30	37	29	23	30	16
37	26	37	28	44	43	10	8
43	37	29	44	11	19	2	7
41	43	23	43	19	39	6	9
37	32	30	10	2	6	2	12
33	18	16	8	7	9	12	11

(d)-Hadamard

0	23	11	22	5	22	10	22
23	34	30	12	38	24	34	22
11	30	31	24	22	29	28	26
22	12	24	13	28	21	27	13
5	38	22	28	11	32	17	30
22	24	29	21	32	22	33	24
10	34	28	27	17	33	24	30
22	22	26	13	30	24	30	17

Table 1. Figure of Merit of the bands of different (64 band) decompositions when the image has to survive JPEG-25 or SPIHT 0.5 bpp. (a) Hartley, (b) DCT, (c) Uniform Subband and (d) Hadamard.

Figure 8. The Energy Compaction Scale

8. THE IDEAL DECOMPOSITION

Note that in Eqn. 5.17, if $\alpha = 0.5$, the capacity of *each* sub-channel of a decomposition is given by

$$\boldsymbol{C}_{h_j} = \log_2(1 + \frac{K\sigma_{\mathrm{ig}_j}}{\sigma^2_{\mathrm{ig}_j} + \sigma^2_{p_j}}) \tag{8.19}$$

In order to maximize $\boldsymbol{C}_{h_j}$ it is enough to maximize $t = \frac{\sigma_{\mathrm{ig}_j}}{\sigma^2_{\mathrm{ig}_j} + \sigma^2_{p_j}}$. It can be easily seen, that t (and hence $\boldsymbol{C}_{h_j}$) is maximized when $\sigma^2_{\mathrm{ig}_j} = \sigma^2_{p_j}$. The ideal decomposition would be the one which results in image noise variances close to the processing noise variances in the maximum number of sub-bands. Note that σ_i and σ_p cannot be made equal in *all* sub-bands since typically $\sum_j \sigma^2_{i_j} >> \sum_j \sigma^2_{p_j}$. It should also be noted, that a decomposition so obtained would perform as expected only if we are able to assume the same model for the relationship between the coefficient variance and the visual threshold. Therefore, the search for such a decomposition may not be simple, and is a topic of current research.

9. CONCLUSIONS

We have presented an information-theoretic approach to estimate the number of bits that can be hidden in still images. We argue why a decomposition of an image into many frequency bands might enhance the number of bits that can be hidden, and this theoretical claim is supported by simulations. We report the achievable capacities for different decompositions like DCT, Hartley, Hadamard and subband transforms and conclude that the choice of the transform should depend on the robustness required. Transforms with poorer energy compaction properties (like Hartley or Hadamard) would in general be preferable to high energy compaction transforms (DCT or Wavelets) for *typical robustness requirements.*

Figure 8 shows the position of various transforms in the *scale of energy compaction.* To the left is the identity transform with no energy compaction. At extreme right is the Karhunen-Leove Transform (KLT), the best energy compaction transform. KLT would yield the best results if processing noise is very low. But as processing noise increases (if we desire greater robustness), we should have to move more and more to the left of the scale to choose a transform.

Note that we evaluate processing noise by measuring the correlation between the image components before and after compression. By this, we implicitly assume that the message signal (signature) is affected to the same extent as the image coefficients themselves by the compressor / decompressor. In a more practical scenario, this might not be true. As in general, the signature will be a random sequence, the compressor / decompressor would suppress the signature to a greater extent than it suppresses the image coefficients. In fact, an ideal compressor should completely eliminate the signature while still retaining significant information about the image. So this may imply higher processing noise than the values obtained from our simulations. At the least, this would imply significantly reduced degrees of freedom for the choice of the signature.

Also, we could decompose the high-image-noise and low-processing-noise low-frequency bands further, resulting in few bands with much lower image noise and low processing noise. This would imply more capacity for the low frequency bands than our simulations predict. In a companion paper in this volume,[9] we propose a non-linear signal addition and detection method, where we use such a decomposition to suppress image noise in the low frequency bands.

REFERENCES

1. I.J. Cox, J. Kilian, F.T. Leighton, and T.G. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transactions on Image Processing, **6** (12) pp 1673-1687, 1997.
2. W.Zeng, B. Liu, "On Resolving Rightful Ownerships of Digital Images by Invisible Watermarks", Proceedings of ICASSP ICIP-97, vol **1** pp 552-555.
3. S. Craver, N. Memon, B-L. Yeo, M. Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships", Proc. IS & T/ SPIE Electronic Imaging: Human Vision and Electronic Imaging, Vol **3022**, pp 310-321, Feb. 97.
4. J. R. Smith and B. O. Comiskey, "Modulation and Information Hiding in Images", Workshop on Information Hiding, University of Cambridge, UK, 30 May - 01 Jun. 1996.
5. T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, John-Wiley and Sons Inc, 1991.
6. Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd Edition, McGraw Hill Inc. 1991.
7. A.Said and W.A.Pearlman, "A New Fast and Efficient Implementation of an Image Codec Based on Set Partitioning in Hierarchical Trees", IEEE Transactions on Circuits and Systems for Video Technology, Volume **6**, pp. 243-250, June 1996.
8. A. N. Akansu, R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands and Wavelets*. Academic Press Inc., 1992.
9. M. Ramkumar, A. N. Akansu, "A Robust Scheme for Oblivious Detection of Watermarks / Data Hiding in Still Images". To be presented in SPIE's Symposium on Voice, Video and Data Communication (VV-06), Boston, MA, 2-5 November 1998.