Multimedia Data Hiding Process

ZORAN BOJKOVIC (1), ALEKSANDAR ROGAN (2)
(1) Faculty of Transport and Traffic Engineering, University of Belgrade
Vojvode Stepe 305, 11000 Belgrade
SERBIA

(2) Public Enterprise of PTT Communications "Srbija" - Operational unit "Zemun"
Glavna 8 - 11080 Zemun, Belgrade
SERBIA

Abstract: - The objective of this paper is to evaluate and assess the potential of multimedia Data hiding. The adaptive least significant bit LSB concept is presented to pixels of different grey levels. Data hiding in the presence of just noticeable difference JND models is taken into account, too. Aggressive compression and data hiding are invoked and analyzed. Assuming the same structure using JND models for data hiding and perceptual coding, we see that at best the data hiding capacity is equal to the increased storage requirements of the capacity of the information.

Key-Words: Perceptual coding, least significant bit, multimedia data, watermark embedding, aggressive compression, aggressive data hiding

1 Introduction

In recent years, data hiding within a multimedia signal has received growing interest due to its potential for signal captioning, maintaining audit trails in media commerce and copy protection through development of digital watermarking technology. Data hiding is the general process by which a discrete information stream is merged within media content by imposing imperceptible changes on the original host signal [1]. One of the main obstacles within the data hiding community has been developing a scheme which is based on the fact that minor modifications of the signal representation will be noticeable in the displayed content. These modifications are imposed on the signal in such a way as to reduce the member of information bits required for storage of the content. Human perceptual models are derived to determine the changes on a signal which remain imperceptible. A duality exists between the problems of perceptual coding and data hiding. The former problem attempts to remove irrelevant and redundant information from a signal, while the latter uses the irrelevant information to mask the presence of the hidden data. Thus, the goals of data hiding and perceptual coding can be viewed as being somewhat contradictory. As a result some papers have dealt with integrating perceptual coding with data hiding [2,3,4], while others have investigated the theoretical relationship between both processes [5]. The central topic of the all works cited above is that there must be an appropriate compromise between data hiding and compression to develop a method which performs both of them. Date hiding decreases the overall possible compression ratio and perceptual coding tampers with the hidden information, so that extraction is difficult.

Image data embedding has been a popular research issue in the last years. It concerns mainly about embedding some proprietary data into digital media for the purpose of identification, annotation and message transmission [6]. Applications can often be found in two fields: one is the digital watermarking and the other is hiding of secret data with a last or cover signal.

Starting from the fact than data hiding and compression have contradictory goals, we provide the usual increment threshold, or just noticeable difference JND perceptual model. After invoking communications analogy for data hiding in the second part of the paper, we deal with aggressive compression and data hiding models. Some final remarks conclude the paper.
2 Visual Increment Threshold

The Visual Increment threshold, or just noticeable difference JND, is defined as the amount of light $\Delta B_T$ necessary to add to a visual field of intensity $B$ such that it can be discriminated from the background. In human visual system (HVS) the curve for $\Delta B_T$ versus $B$ can be analytically modeled [7]. At low intensity near the absolute visual threshold (i.e., presence or absence of light intensity detectable under a dark conditions), the value of $\Delta B_T$ is constant. With an increase in intensity $B$, the incremental visual threshold follows the square root law (known as the De Vries - Rose region), i.e., $\Delta B_T \propto \sqrt{B}$. Further ahead, the behavior converges to the well-known Weber's law, i.e., $\Delta B_T \propto B^\gamma$ and finally comes to a saturation region. Obviously, an equal amount of $\Delta B_T$ created at different background intensity $B$ does not result in an equal perceivable change to the HVS.

We can approximate the visual incremental threshold curve by piecewise functions. When this is done, the threshold values in the De Vries - Rose region, the Weber region, and the saturation region are defined by [8]

\[
\log \Delta B_T = \log (K_1 \cdot B) \quad \text{(Weber)} \tag{1}
\]

\[
\log \Delta B_T = \log (K_2 \cdot \sqrt{B}) \quad \text{(De Vries - Rose)} \tag{2}
\]

\[
\log \Delta B_T = \log (K_3 \cdot B^2) \quad \text{(Saturation)} \tag{3}
\]

respectively, while $K_1$, $K_2$, and $K_3$ are the constants of proportionality. According to the Weber's law the value of $\frac{\Delta B_T}{B}$ remains fairly constant with a value approximately equal to $\beta$ of $\left(\frac{\Delta B}{B}\right)_{\text{max}}$. The maximum value of $\frac{\Delta B}{B}$ over the entire image can be obtained from:

\[
K_1 = \frac{\Delta B_T}{B} = \beta \left(\frac{\Delta B}{B}\right)_{\text{max}} \tag{4}
\]

Relations of the incremental threshold $\Delta B_T$ vs. reference intensity $B$ in different regions are shown in Fig.1. The following equality should be satisfied at the interception of the Weber and De Vries - Rose regions. Thus, we have

\[
\log x_2 + \log K_1 = \frac{1}{2} \log x_2 + \log K_2 \tag{5}
\]

and obtain

\[
K_2 = K_1 \cdot \sqrt{x_2} \tag{6}
\]

Similarly, the Weber and the Saturations regions intercept at the point $x_3$ and we obtain

\[
K_3 = \frac{K_1}{x_3} \tag{7}
\]

Thus, the minimum amount of the incremental thresholds in three different regions can be completed as

\[
\Delta B_r = K_2 \cdot \sqrt{B} \tag{8}
\]

\[
\Delta B_r = \sqrt{B} \cdot \beta \cdot \left(\frac{\Delta B}{B}\right)_{\text{max}} \cdot \sqrt{x_2}, \text{ for } x_1 \leq B < x_2 \tag{9}
\]

\[
\Delta B_r = K_3 B^2 = B^2 \cdot \beta \cdot \left(\frac{\Delta B}{B}\right)_{\text{max}} \cdot \frac{1}{x_3}, \text{ for } x_3 < B \tag{10}
\]

The absolute visual threshold at low intensities signals

\[
\Delta B_T = \sqrt{x_1} \cdot x_2 \cdot \beta \left(\frac{\Delta B}{B}\right)_{\text{max}}, \text{ for } 0 \leq B < x_i \tag{12}
\]

A typical set of parameters is considered as $\beta \simeq 0.04$, $\left(\frac{\Delta B}{B}\right)_{\text{max}} = 2.05$, and $x_i = \alpha_i \cdot 255$

where $\alpha_1$, $\alpha_2$, and $\alpha_3$ equal to 0.1, 0.3 and 0.9 respectively, to complete $\Delta B_T$ for each possible reference intensity $B$. These $\Delta B_T$ values are considered as a threshold function $C(g)$, $g=0, ..., 255$, to be referred in subsequent processing for the determination of adaptive number of the least significant bit (LSB).
3 Adaptive Least Significant Bit Method

Least significant bit method hides data in a fixed number of LSBs of an image pixel. The more the LSBs are changed, the more distortion we get. The number of LSBs for data embedding is pixel-dependent according to the above discussion. In order to adaptively modify grey levels within different tolerances of HVS sensitivity to contrast variation, we invoke an LSB mapping function \( N(g) \). If gives the number of LSBs that can be embedded in each possible gray level \( g \). Let the pixel value after embedding be denoted by \( g' \). The condition of no overhead for this adaptation is that the \( N(g) \) values before and after pixel modification should remain unchanged. It means that \( N(g') = N(g) \). The concept of self-contained member of LSBs for data embedding and extraction is shown in Fig.2.

The grey level difference \( |g' - g| \) should be subject to the constraint of \( C(g) \) so that it is little perceivable. Basically, the curve of \( N(g) \) is made piecewise step, i.e.

\[
N(g) = \begin{cases} 
1, & 0 \leq g < k_1 \\
2, & k_1 \leq g < k_2 \\
3, & k_2 \leq g < k_3 \\
4, & k_3 \leq g < 256 
\end{cases}
\]  

A value larger than 4 would not be considered since modifications of too many LSBs is likely to results in perceivable change in image contrast. The key procedure to determine the decision boundaries \( k_1 \sim k_2 \) is as follows

- Set \( N(g) \) to be the largest value \( n_g \) which satisfies \( 2^{n_g} - 1 \leq C(g) \).
- Examine \( N(g) \) to find \( k_1, k_2 \) and \( k_3 \) where the values change.

4 Communications and Data Hiding

Communicating the hidden signal information is likened to transmission of the signal through an associated communications channel as is shown in Fig.3. Embedding the signal is equivalent to channel coding and extraction of the hidden information serves the same purpose as a communications receiver. For most data hiding applications, the only potential source of manipulation after embedding is perceptual coding. In this case, the process of lossy compression characterizes the associated communications channel for the hidden data.

The particular measure we are concerned with is that of transmission capacity. Employing the just
noticeable difference JND, we can treat the problem as an information theoretic one. Communication channel is considered to be confused of smaller sub-channels denoted $C_i$, for $i = 1, 2, ..., M$. Consider the discrete signal $f(i)$ transformed with $T$ to produce the set of coefficients $F(u)$. Each $F(u)$ will have an associated JND, denoted by $J'(u)$, such that we can form $F(u)$ as follows

\[ F'(u) = F(u) + \beta(u)J'(u) \]  (14)

where $\beta(u)$ is any signal with coefficient between the values -1 and 1. The inverse transform $T^{-1}$ of $F'(u)$ produces the signal $f(i)$ which is guaranteed to be perceptually identical to $f(i)$. The idea is to make the JND values as large as possible to exploit the masking characteristic of a broad class of signals. We assume the individual perceptual models used for data hiding and compression are conservative. Specifically, if the data hiding algorithm is restricted to making changes to the coefficient $F(u)$ below or equal in magnitude to $\alpha(u)$, then the compression algorithm must have an effective JND for quantization of $J(u) = J'(u) - \alpha(u)$ to be both efficient and yet cause no visual distortions.

Assume that the coefficient $F(u)$ are granted into disjoint sets $G_i$, such that if $F(v) \in G_i$, then

\[ \frac{\alpha(v)}{J(v)} = \epsilon_i \]  (15)

for some positive value $\epsilon_i$, which we call the relative perceptual efficiency. The values of $\epsilon_i$ do not necessarily have to be distinct for each $i$. The number of elements $N$ in $G_i$ is sufficiently large that a perceptual length channel code may be used to transmit watermark information. Let $W(u)$ represent the signal change in $F(u)$ to embed hidden data. After compression, assuming no interference from the host signal, the received signal is

\[ \hat{W}(u) = W(u) + Q(u) \]  (16)

where $Q(u)$ is additive uniformly distributed noise and is assumed to be independent of $W(u)$. In what follows, we will drop the argument $u$.

Consider $|W| \leq \alpha$. The probability density function (pdf) of $Q$ is given by

\[ p_Q(q) = \begin{cases} \frac{1}{2J}, & \text{for } |q| \leq J \\ 0, & \text{otherwise} \end{cases} \]  (17)

Tacking into account the independence of $W$ and $Q$, it will be

\[ p_W(\xi) = p_W(\xi) \cdot p_Q(\xi) \]  (18)

where we use the notation that $p_W$ is the pdf of random variable $X \in (\hat{W}, W, Q)$.

### 5 Aggressive Compression

Aggressive compression is defined as a case for which $\alpha < J$, or equivalently $\epsilon_i < 1$. In this case, equation (18) is reduced to

\[ p_W(\xi) = \begin{cases} 0, & \text{for } \xi \leq -\alpha - J \\ \frac{1}{2J} \cdot p_W(\xi + J), & \text{for } -\alpha - J < \xi \leq \alpha - J \\ \frac{1}{2J} \cdot [1 - p_W(\xi + J)], & \text{for } -\alpha + K < \xi \leq \epsilon + J \\ 0, & \text{for } \xi > \alpha + J \end{cases} \]  (19)

where $p_W(\xi)$ is the cumulative distribution function (cdf) of $W$, i.e.,

\[ p_W(\xi) = \int_{-\infty}^{\xi} p_W(\eta) d\eta \]  (20)

The data hiding capacity of the associated subchannel $C_i$ is defined as

\[ C_i = \max_{p_W(\xi)} I(\hat{W}, W) \]  (21)

where $p_W(\xi)$ is the pdf of $W$, and $I(\cdot, \cdot)$ is the mutual information between the two argument distributions [9].

Taking the derivative of $I(\hat{W}, W)$ with respect to $p_W(\xi)$ and equating to zero, if can be find that the capacity is achieved for

\[ p_W(\xi) = \begin{cases} 0, & \text{for } \xi \leq -\alpha \\ \frac{1}{2}, & \text{for } -\alpha < \xi \leq \alpha \\ 1, & \text{for } \xi > \epsilon \end{cases} \]  (22)

or equivalently,

\[ p_W(\xi) = \frac{1}{2} \delta(\xi + \alpha) + \frac{1}{2} \delta(\xi - \alpha) \]  (23)

where $\delta(\xi)$ is the Dirac delta function. This means that to achieve capacity, $W$ has a discrete binary uniform distribution. Then the capacity of the subchannel is given by

\[ C_i = H(\hat{W}) - H(Q) \]  (24)
\[ C_i = \left[ \frac{\alpha}{J} + \log (2J) \right] \]  \hspace{1cm} (25)

\[ C_i = \frac{\alpha}{J} \]  \hspace{1cm} (26)

\[ C_i = \epsilon_i \]  \hspace{1cm} (27)

where \( H(*) \) is the entropy of the argument random variable.

Assuming that the signal and noise are independent for each sub-channel \( C_i \), the overall data hiding channel capacity is given by

\[ C = \sum_{i=1}^{M} C_i = \sum_{i=1}^{M} \epsilon_i \]  \hspace{1cm} (28)

The compression sacrifice \( CS \) due to data hiding is defined as the number of additional bits required for storage of the signal because some of the perceptual masking properties are used for data hiding. For the \( u \)-th coefficient it is given by \( \log_2 \left( \frac{\alpha + J}{J} \right) \). Thus, for sub-channel \( C_i \), the bit rate sacrifice is \( N \log_2 (\epsilon_i + 1) \). This is related to the sub-channel capacities by

\[ CS = N \sum_{i=1}^{M} \log_2 (C_i + 1) \]  \hspace{1cm} (29)

where \( N \) is the number of coefficient comprising each sub-channel.

### 6 Aggressive Data Hiding

For dominant data hiding, \( \epsilon_i \geq 1 \). Similarly, the capacity \( C \) of each sub-channel is bounded as follows

\[ \log_2 \left( \left\lfloor \epsilon_i \right\rfloor + 1 \right) \leq C_i \leq \log_2 \left( \left\lceil \epsilon_i \right\rceil + 1 \right) \]  \hspace{1cm} (30)

where \( \left\lfloor \cdot \right\rfloor \) and \( \left\lceil \cdot \right\rceil \) are the floor and ceiling operators, respectively.

The probability distribution of \( W \) which achieves capacity is discrete and uniformly distributed. Thus, we have

\[ \sum_{i=1}^{M} \log_2 \left( \left\lfloor \epsilon_i \right\rfloor + 1 \right) \leq C \leq \sum_{i=1}^{M} \log_2 \left( \left\lceil \epsilon_i \right\rceil + 1 \right) + 1 \]  \hspace{1cm} (31)

In the case that \( \epsilon_i \) is a natural number it will be

\[ CS = NC \]  \hspace{1cm} (32)

It means that the bit rate sacrificed for more efficient storage is equal to the data hiding capacity bit rate multiplied with the number of coefficients in each sub-channel.

### 7 Watermark Embedding and Perceptual coding

Due to using the similar and transforms for data hiding and compression, the general trend is to embed the watermark in the same domain as for performing perceptual coding [10].

Data hiding of the information signal \( W \) into the digital multimedia signal \( f \) occurs in the transform \( T \) domain. Specifically, the data \( W \) is hidden in the discrete coefficients \( F(u) \) produced by applying the invertible \( T \) on \( f \). The new coefficient from the embedding process \( \hat{F}(u) \) are transformed using the inverse of \( T \) to produce the output of the data hiding process denoted \( \hat{f} \). For each coefficient \( -F(u) \), the signal change due to watermark embedding to produce \( \hat{F}(u) \) does not exceed \( \alpha(u) \) which is below the JND threshold \( J(u) \) for that coefficient.

Perceptual coding of a signal \( \hat{f} \) occurs in the same domain as data hiding, after signal embedding. The embedded signal \( \hat{f} \) is transformed with \( T \) to produce coefficients which are then quantized to reduce the signal storage requirements. It there was no data hiding, \( \alpha(u) = 0 \) for all \( u \). The lossy compression would be equivalent to standard JND perceptual coding using \( J(u) \). The only source of error on the extracted information is due to lossy compression. The host signal does not provide any interference to the hidden data.

### 8 Concluding Remarks

For embedding multimedia information (audio, image, video, or text compressed or non-compressed) into a host image subject to the constraint of human visual perception, the classical least significant bit method LSB can be applied, but with adaptive numbers of LSBs for different grey levels pixels. The number of LSBs for embedding can be figured out from the modified pixel value without extra overhead.
Assuming the same structure using just noticeable difference JND models for data hiding and perceptual coding, it can be seen that the best the data hiding capacity is equal to the increased storage requirements of the information.

We can also see that the gain in hiding ability is at most what we loose in compression efficiency. This motivates future investigation to use different transforms for both tasks. The used complementary domains for the hiding and compression process may allow to hide information without sacrificing compression efficiency.

References


