A JOINT SOURCE-CHANNEL CODING SCHEME FOR IMAGE-IN-IMAGE DATA HIDING

K. Solanki, O. Dabeer, B. S. Manjunath, U. Madhow, and S. Chandrasekaran

Dept. of Electrical and Computer Engineering
University of California at Santa Barbara
Santa Barbara, CA 93106
Email: {solanki, onkar, manj, madhow, shiv}@ece.ucsb.edu

ABSTRACT

We consider the problem of hiding images in images. In addition to the usual design constraints such as imperceptible host degradation and robustness in presence of variety of attacks, we impose the condition that the quality of the recovered signature image should be better if the attack is milder. We present a simple hybrid analog-digital hiding technique for this purpose. The signature image is compressed efficiently (using JPEG) into a sequence of bits, which is hidden using a previously proposed digital hiding scheme. The residual error between the original and compressed signature image is then hidden using an analog hiding scheme. The results show (perceptual as well as mean-square error) improvement as the attack becomes milder.

1. INTRODUCTION

Driven by applications such as steganography, digital watermarking, addition of meta-content, and document authentication, there has been a growing body of work in data hiding (see, for example, [1, 2, 3, 4, 5, 6], and references therein). We consider the problem of image-in-image hiding in this paper, where, the basic design criteria are as follows: (a) the degradation to the host image is imperceptible, (b) it should be possible to recover the hidden, or signature, image under a variety of attacks, and (c) the quality of the recovered signature image should be better if the attack is milder. In recent work [1, 2, 3, 5, 7], it has been shown that digital data can be effectively hidden in an image so as to satisfy criteria (a) and (b) by hiding in the choice of quantizer for the host data. The main idea is to view the data hiding problem as communication with channel side information ([8, 9, 4]): the channel experienced by the data comprises of the host interference and the attack, and the channel side information is the knowledge of the host. Therefore, recent advances in source coding and channel coding can be leveraged for developing data hiding schemes.

Unfortunately, these schemes do not satisfy the design criterion (c) - they exhibit the threshold effect: if the actual attack is more severe than the attack the scheme was designed for, there is a catastrophic failure in recovering the hidden image, while if the actual attack is less severe, then we are still stuck with the design attack image quality. In practice, the attack level is seldom known apriori, and ideally, we would like a scheme that results in graceful improvement and degradation in the image quality with less and more severe attacks respectively. Such schemes require *joint* source-channel coding, which has been studied for the Gaussian channel in [10, 11, 12]. However, to the best of our knowledge, such schemes have not been studied for the data hiding channel. In this paper, we present a hybrid digital-analog (joint source-channel) coding scheme for image-in-image hiding. It leverages an earlier digital scheme based on image-adaptive criteria and turbo-like repeat-accumulate (RA) codes [2, 3], and involves the transmission of the analog residue using a new method, which is similar in flavor to the quantization index modulation commonly used in digital schemes. At the decoder, we focus on JPEG attacks. The proposed scheme shows (perceptual as well as mean-square error) improvement over the purely digital scheme in [2, 3] as the level of the JPEG compression attack decreases.

The rest of the paper is organized as follows. In Section 2, we provide a brief background of joint source-channel coding. In Section 3, we describe our method for transmitting the analog residue and derive the minimum mean-square error estimator (MMSE) for the analog signature under uniform quantization attack. We assume that the quantization matrix of the JPEG attack is known to the decoder. In Section 4, we describe our hybrid digital-analog scheme and present the results. We present the conclusions in Section 5.

2. JOINT SOURCE-CHANNEL HIDING

To provide a background for joint source-channel coding, we first briefly consider some fundamental limits for the Gaussian data hiding channel ([8, 9]). Consider an i.i.d. Gaussian signature source with zero mean and variance σ^2, which has to be embedded in a Gaussian host. The hider is at most allowed to introduce a mean-square error D_1 per host symbol. Further, we assume a Gaussian attack (that simply adds i.i.d. Gaussian noise), which introduces an additional distortion of at most D_2 per host symbol. In general, the host and the signature have different sizes, and so we assume that ρ channel uses per source symbol are allowed. At the receiver, we are interested in recovering the signature with distortion D_3 per signature symbol. From the information capacity results from [8], and rate distortion theory ([13]), we can easily deduce that,

$$D_3 \geq \frac{\sigma^2}{\left(1+\frac{D_1}{D_2}\right)^{\rho}} =: D_{min}. \qquad (1)$$

Given D_1, D_2 and ρ, the smallest feasible distortion above can be approached in principle by separate source and channel coding. Unfortunately, such schemes have the drawback that even if the Gaussian attack channel introduces a distortion less that D_2, we suffer distortion D_{min}, though in principle we can have smaller

This research was supported in part by a grant from ONR # N00014-01-1-0380.

distortion. One of the goals of joint source-channel coding is to provide improvement for less severe attacks. For the Gaussian channel (that is, the host is absent), a number of joint source-channel coding methods have been proposed. In [10], codes based on chaotic systems have been proposed, which recently were shown to have optimal scaling properties in the high signal-to-noise regime in [12]. In [11, 14], hybrid digital-analog codes have been proposed. However, for the data hiding channel, joint source-channel codes have not been studied so far and a number of issues are open. In this paper, we exhibit a practical hybrid digital-analog scheme for image-in-image hiding, which is similar to the scheme proposed in [14] for the Gaussian channel. The idea is to compress the signature image efficiently into a sequence of bits, which is hidden using a previously proposed digital hiding scheme [3]. The residual error between the original and compressed signature image is then hidden using an analog hiding scheme (proposed in Section 3). With practical issues in mind, we focus our attention to JPEG compression attacks instead of the Gaussian attack. We chose to develop a hybrid digital-analog scheme for the following purposes.

1. It allows us to exploit advantages of the digital scheme in [2, 3], which hides high volume of data using image-adaptive criteria and turbo-like codes, and is also robust against a variety of attacks.
2. Due to the limited dynamic range of the analog residue, it is feasible to send them reliably over a limited number of host symbols.

In the next section, we first describe the analog part of our scheme in detail. The actual scheme and simulation results are given in Section 4.

3. HIDING ANALOG INFORMATION

In this section, we propose a strategy to hide an analog number into a host sample. The hiding strategy involves quantization of the host followed by replacing the residue with the appropriately scaled source and is given in Section 3.1. The MMSE decoder is derived in Section 3.2.

3.1. Hiding using scalar quantization of the host

To hide an analog number m into a host sample h, we first quantize the host h using a quantizer of step size Δ, and then replace the residue with the source m, which has been companded or scaled to lie in the interval $(0, \Delta)$. Let us consider an example where $\Delta = 1$ and the host symbol is, say, 6.235. We want to send a source symbol whose value is 0.729 (a real number $\in (0, \Delta)$) through the hiding channel. The encoder first determines that the host symbol lies between 6 and 7 (an interval $(n\Delta, (n+1)\Delta)$), then it sends the source symbol directly within that interval, i.e., it just sends 6.729. In practice, we use a hiding strategy that always *measures* the message m from an even reconstruction point of the host. This is done to avoid catastrophic error when a hidden coefficient switches to a different integer interval as a result of attack. Thus, the symbol y to be sent for hiding a message m into a host symbol h is given by,

$$\begin{aligned} y &= \Delta(\lfloor h/\Delta \rfloor) + m, \text{ if } \lfloor h/\Delta \rfloor \text{ is even,} \\ &= \Delta(\lfloor h/\Delta \rfloor + 1) - m, \text{ if } \lfloor h/\Delta \rfloor \text{ is odd.} \end{aligned} \tag{2}$$

Here, $\lfloor \cdot \rfloor$ denotes the `floor` operation (defined as the largest integer smaller than or equal to the given number).

3.2. JPEG attacks and MMSE decoding

The JPEG compression performs uniform quantization of the Discrete Cosine Transform (DCT) coefficients of 8×8 blocks of the image. Hence we derive the MMSE decoder for the above hiding scheme under uniform quantization attack, when the reconstruction points of the attack quantizer are known to the decoder, but not to the encoder. In this section, we use bold italics to represent random variables; their realizations are denoted by corresponding italic letters.

We consider the case of hiding a uniform random variable $\boldsymbol{m} \sim U[0,1]$ using (2) into an independent host coefficient $\boldsymbol{h}$ to obtain $\boldsymbol{y}$. In practice, even if $\boldsymbol{m}$ is not $U[0,1]$, it can be transformed in to an uniform random variable by applying the inverse of its distribution function. Without loss of generality, we assume $\Delta = 1$. In this analysis, we restrict our attention only to attacks with quantization interval less than or equal to the design interval. Note that, in practice, the design interval will be an entry in the design JPEG quantization matrix, which will be chosen to be the worst case attack. Denoting the attack quantization interval by $\delta \leq 1$, the received symbol $\boldsymbol{z} = Q(\boldsymbol{y})$, where $Q(\cdot)$ denotes the uniform quantization with an interval δ, and with zero as one of the reconstruction points. Note that all JPEG quantizers have zero as one of its reconstruction points. Thus, $\boldsymbol{z} \in \{..., -2\delta, -\delta, 0, \delta, 2\delta, ...\}$. The MMSE decoder is simply the conditional expectation $E[\boldsymbol{m}|\boldsymbol{z} = z]$. In the following, we consider various cases depending upon z, and find the conditional expectation by identifying the conditional density of $\boldsymbol{m}$ given $\boldsymbol{z} = z$.

If $z = a\delta$ is received, then y necessarily lies in the interval $[(a - 1/2)\delta, (a + 1/2)\delta)$, which we call its *ambiguity* interval. Let us consider the integer interval in which z is received, say $[n, n+1)$. There are three possibilities with the ambiguity interval: (i) No crossing: The ambiguity interval for y does not cross into another integer interval, that is,

$$z - \frac{\delta}{2} \geq n \text{ and } z + \frac{\delta}{2} < n + 1. \tag{3}$$

(ii) Even crossing: The ambiguity interval crosses an even integer, that is,

$$\begin{aligned} z - \frac{\delta}{2} &< n \text{ and } n \text{ is even, or,} \\ z + \frac{\delta}{2} &\geq n + 1 \text{ and } (n+1) \text{ is even.} \end{aligned}$$

(iii) Odd crossing: The ambiguity interval crosses an odd integer, that is,

$$\begin{aligned} z - \frac{\delta}{2} &< n \text{ and } n \text{ is odd, or,} \\ z + \frac{\delta}{2} &\geq n + 1 \text{ and } (n+1) \text{ is odd.} \end{aligned}$$

Now we proceed to find the MMSE estimates of the message $\boldsymbol{m}$ for all the three cases.
(i) No crossing: In this case,

$$f_{\boldsymbol{m}|\boldsymbol{z}}(m|z) = U\left[(a - 1/2)\delta, (a + 1/2)\delta\right).$$

The corresponding MMSE estimate is,

$$\hat{m} = \begin{cases} z - n & \text{if } n \text{ is even,} \\ (n+1) - z & \text{if } n \text{ is odd.} \end{cases} \tag{4}$$

(ii) Even Crossing: As mentioned above there could be two cases for even crossing, each involving either n or $(n+1)$ being even. The analysis is similar in both the cases and hence we just consider the first case (n even). Let us define $R_1 = n - (z - \delta/2)$ and $R_2 = (z + \delta/2) - n$ as the distances between the even crossing point n, and, the lower and upper points of the ambiguity interval respectively. Note that $R_1 + R_2 = \delta$. Defining the events $A := \{\boldsymbol{y} \in [n - R_1, n)\}$ and $B := \{\boldsymbol{y} \in [n, n + R_2)\}$, we have,

$$f_{\boldsymbol{m}|\boldsymbol{z}}(m|z) = f_{\boldsymbol{m}|\boldsymbol{z},A}(m|z, A) \cdot P(A|z) + f_{\boldsymbol{m}|\boldsymbol{z},B}(m|z, B) \cdot P(B|z)$$

where,

$$\begin{aligned} P(A|z) &= P(\lfloor \boldsymbol{h} \rfloor = (n-1), \boldsymbol{m} \in [0, R_1] | \boldsymbol{z} = z) \\ &= \frac{P(\lfloor \boldsymbol{h} \rfloor = (n-1), \boldsymbol{m} \in [0, R_1], \boldsymbol{z} = z)}{P(\boldsymbol{z} = z)} \\ &= \frac{P(\lfloor \boldsymbol{h} \rfloor = (n-1)) \cdot P(\boldsymbol{m} \in [0, R_1])}{P(\boldsymbol{z} = z)} \\ &= \frac{P(\lfloor \boldsymbol{h} \rfloor = (n-1)) \cdot R_1}{P(\boldsymbol{z} = z)}. \end{aligned} \tag{5}$$

Similarly,

$$P(B|z) = \frac{P(\lfloor \boldsymbol{h} \rfloor = n) \cdot R_2}{P(\boldsymbol{z} = z)} \tag{6}$$

where,

$$P(\boldsymbol{z} = z) = P(\lfloor \boldsymbol{h} \rfloor = (n-1)) \cdot R_1 + P(\lfloor \boldsymbol{h} \rfloor = n) \cdot R_2.$$

Note that, for a slowly varying host distribution, we have, $P(\lfloor h \rfloor = (n-1)) \approx P(\lfloor h \rfloor = n)$, so that, (5) and (6) can be approximated as $P(A|z) = R_1/\delta$, and $P(B|z) = R_2/\delta$.

Since the event $A \cap \{\boldsymbol{z} = z\} = \{\boldsymbol{m} \in [0, R_1)\}$, we have $f_{\boldsymbol{m}|\boldsymbol{z},A}(m|z, A) = U[0, R_1]$. Hence, the MMSE estimate is,

$$\hat{m} = \frac{R_1}{2} P(A|z) + \frac{R_2}{2} P(B|z).$$

Again, for a slowly varying host distribution, after some simplifications, we get,

$$\hat{m} = \frac{\delta}{2} - \frac{R_1 R_2}{\delta}. \tag{7}$$

(iii) Odd crossing: Following the analysis of the even case, define R_1 and R_2 as distances between the crossing point and lower and upper points of the ambiguity interval respectively. Here, we get the MMSE estimate for the general case as,

$$\hat{m} = \frac{2 - R_1}{2} P(A|z) + \frac{2 - R_2}{2} P(B|z)$$

and for the slowly varying host distribution, we get,

$$\hat{m} = 1 - \Big(\frac{\delta}{2} - \frac{R_1 R_2}{\delta}\Big). \tag{8}$$

Hence, we have the MMSE estimate for all the cases which can be used for decoding when decoder knows the JPEG compression quantization matrix.

4. IMAGE-IN-IMAGE HIDING

In this section we describe the actual implementation of the entire system for image-in-image hiding. The encoding process can be divided into following parts.

Processing the signature image: This step involves separating the signature image into digital and analog parts. The image is compressed using JPEG to generate a bitstream, which constitutes the digital part. The analog part is obtained by computing the residual errors of pre-selected DCT coefficients after the quantization based on design *signature* quantization matrix. Note that, the design quality factor, and the number of analog residues chosen to send, are predetermined at the design stage.

Allocating the channels: Here, we allocate the host coefficients (i.e., channel) for the digital and analog parts respectively. A few low frequency coefficients (other than the DC coefficient) of each 8×8 host block are reserved for the analog channel. Remaining low and/or mid frequency coefficients are dedicated to the digital channel. Thus the decoder would know where to look for analog and digital data respectively.

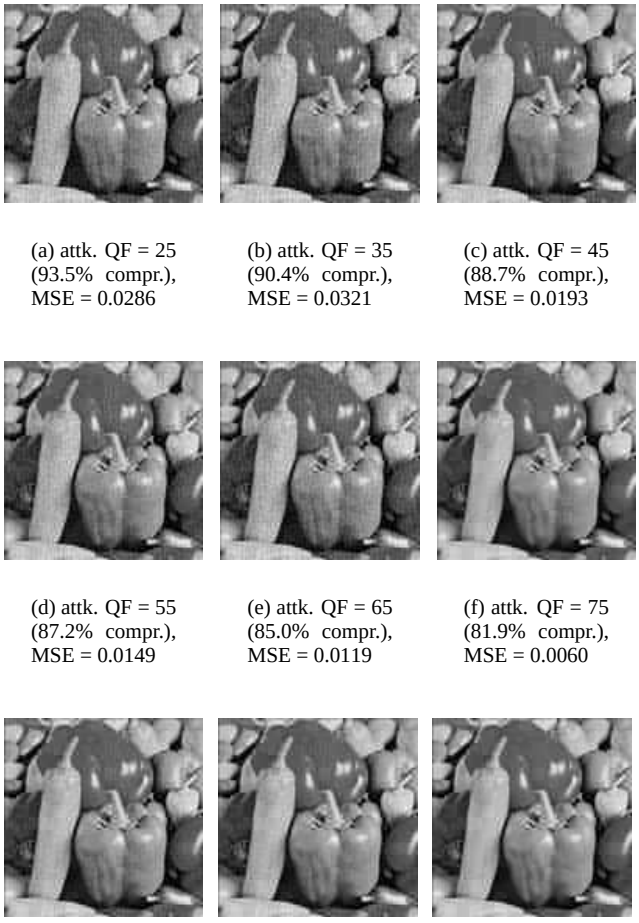
Hiding the digital part: The digital bitstream is hidden into its allocated channel using the RA-coded Selectively Embedding in Coefficients (SEC) scheme of [2, 3]. The bitstream to be hidden is coded using turbo-like RA code at a low rate. This coded bitstream is hidden into the host coefficients such that a code symbol is *erased at the encoder*, if the floor of its magnitude is smaller than or equal to a predetermined integer threshold. The decoder uses the same threshold criteria to estimate the erasure locations. The RA code rate is designed in such a way that one can also deal with the additional errors and erasures due to attack.

Hiding the analog part: The analog residues of selected low frequency coefficients are sent through its allocated channel using the hiding scheme of Section 3. Since the residue always lies in $[0, \Delta_{sig})$, where Δ_{sig} is specified by the design quantizer, we simply scale it to lie in $[0, 1)$.

The decoder decodes the analog and digital parts separately and adds them together to give an estimate of the sent signature image. The decoding of the analog part is done using the knowledge of attack δ, and assuming a slowly varying host distribution (Section 3.2). The digital part is iteratively decoded using sum-product algorithm. Now we present two example implementations to show that there is an improvement in perceptual quality as well as the mean-squared error (MSE) for the received signature image as the attack becomes milder. Note that though we present two specific examples here, the scheme is applicable for any image-in-image hiding scenario.

Example 1: We hide a 128×128 image into a 512×512 image, with the design quality factor of 25. Figure 1 shows the recovered signature images when the host image undergoes JPEG compression at varying levels, starting from the worst case QF of 25. The signature image is JPEG compressed at QF = 10 to form the digital part and the residues of 16 low frequency coefficients make up the analog part. We use one coefficient from each 8×8 host block for transmitting the analog data. 34 coefficients constitute the *digital channel*.

Example 2: A 256×256 image is hidden with a design QF of 50. Table 1 shows the MSE of the received image after varying levels of JPEG compression. The signature image is JPEG compressed at QF=18, and residues of 12 low frequency coefficients constitute the analog part. 3 coefficients per host block are used for sending analog residue and another 32 coefficients form the candidate embedding band for the digital data.



(a) attk. QF = 25 (93.5% compr.), MSE = 0.0286

(b) attk. QF = 35 (90.4% compr.), MSE = 0.0321

(c) attk. QF = 45 (88.7% compr.), MSE = 0.0193

(d) attk. QF = 55 (87.2% compr.), MSE = 0.0149

(e) attk. QF = 65 (85.0% compr.), MSE = 0.0119

(f) attk. QF = 75 (81.9% compr.), MSE = 0.0060

(g) attk. QF = 85 (75.8% compr.), MSE = 0.0043

(h) attk. QF = 95 (57.7% compr.), MSE = 0.0025

(i) No attack

Fig. 1. Example 1: Hiding a 128×128 peppers image into a 512×512 harbour image (not shown here). The signature images received after various levels of JPEG compression are shown along with the corresponding observed MSE per coefficient.

Table 1. Example 2: MSE per coefficients for varying levels of attacks. A 256×256 clock image has been hidden in a 512×512 bridge image.

Attk. QF	50	60	70	80	90
compr.	84.2%	81.9%	78.3%	72.5%	60.0%
MSE/coeff.	0.0335	0.0374	0.0266	0.0146	0.0046

5. CONCLUSIONS

In this paper, we demonstrated a simple hybrid digital-analog scheme for image-in-image hiding. As the JPEG attack quality factor increases, we recover the signature image with better quality. While the results show improvement over a purely digital hiding strategy, much more further work remains in exploring the huge space of possible joint source-channel coding strategies for this application.

References

[1] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Info. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[2] K. Solanki, N. Jacobsen, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Robust image-adaptive data hiding based on erasure and error correction," Under review *IEEE Trans. on Image Processing*.

[3] N. Jacobsen, K. Solanki, S. Chandrasekaran, U. Madhow, and B. S. Manjunath, "Image adaptive high volume data hiding based on scalar quantization," in *Proc. IEEE Military Comm. Conf. (MILCOM)*, Oct. 2002.

[4] J. Chou and K. Ramachandran, "Robust turbo-based data hiding for image and video sources," in *Proc. ICIP*, Oct. 2002.

[5] J. J. Eggers, R. Buml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," To appear, IEEE Trans. on Sig. Pro., Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[6] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1108–1126, 1999.

[7] K. Solanki, N. Jacobsen, S. Chandrasekaran, U. Madhow, and B. S. Manjunath, "High-volume data hiding in images: Introducing perceptual criteria into quantization based embedding," in *Proc. ICASSP*, May 2002.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[9] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. on Info. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[10] B. Chen and G. W. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Trans. on Communications*, vol. 46, no. 7, pp. 881–890, July 1998.

[11] U. Mittal and N. Phamdo, "Hybrid digital-analog joint source-channel codes for broadcasting and robust communications," *IEEE Trans. on Info. Theory*, vol. 48, no. 5, pp. 1082 –1102, May 2002.

[12] V. Vaishampayan and S. I. R. Costa, "Curves on a sphere, shift-map dynamics and error control for continuous alphabet sources," Preprint, Nov. 2002.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[14] M. Skoglund, N. Phamdo, and F. Alajaji, "Design and performace of vq-based hybrid digital-analog joint source-channel codes," *IEEE Trans. on Info. Theory*, vol. 48, no. 3, pp. 1082 –1102, Mar 2002.