14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

JOINT DATA HIDING-SOURCE CODING OF STILL IMAGES

Çagatay Dikici, Khalid Idrissi, and Atilla Baskurt

INSA de Lyon, Laboratoire d'InfoRmatique en Images et Systèmes d'information,
LIRIS, UMR 5205 CNRS, France
{cdikici,kidrissi,abaskurt}@liris.cnrs.fr
http://liris.cnrs.fr

ABSTRACT

There exists a strong duality between Source Coding with side information (known as Distributed Source Coding) and Channel Coding with Side Information (informed data-hiding). Inspired by the system in [1], which is a combination of Data Hiding and Distributed Source Coding scheme, we have extended this system to 2D natural images. Hence a cascade of informed data hiding for a still image is followed by a distributed source coding with a fidelity criterion, given that a noisy version of the input image is available only at the decoder. We used baseline JPEG compression with different quality values for creation of the Side Information, a memoryless quantization for informed data hiding, and LDPC based coset construction for source coding. The preliminary experimental results are given using the theoretical findings of the duality problem.

1. INTRODUCTION

The information theoretical limits of Rate-Distortion function of Source Coding with Side Information(SCSI) [2, 3], and the channel capacity of Channel Coding with Side Information(CCSI) [4, 5] at gaussian case has been clearly understood since 1970s. After the invention of good channel codes like turbo[6] and LDPC[7], these codes have been applied to both Distributed Source Coding[8, 9] and Informed Watermarking[10, 11, 12] problems. Departing from the strong duality between the SCSI and CCSI[13, 14, 15], we propose a data-hiding and compression scheme for still-images using the mixture of these two setup. Our proposed system enables to embed a hidden message M into a host image X, continued with a DSC compression scheme, given that a noisy version Y of the host signal is available only at the decoder. Both the hidden message $\hat{M}$ and the input image $\hat{X}$ is estimated at the decoder.

Since Distributed Source Coding enables to shift encoder complexity to the decoder side, it can be used on several application scenarios for low-power devices, from image and video coding to sensor networks. For example, in the case of encoding correlated observations of low-power sensor networks, Distributed Source Coding principles fit best for achieving power constraints. In informed data hiding or known as blind watermarking system, the watermarked data depends on both host signal and embedded message, and the message can be extracted at the decoder blindly. Data Hiding of Moreover in this system, each sensor can easily hide its own hidden data within the observed signal and be coded with distributed source coding principles. This hidden data could be also served for digital rights management or contains additional information of the sensor like the coordinates of the camera or region of interest information.

In this paper, we look at the informed data-hiding problem of a distributed source coding system shown in Fig. 1. From the nature of Distributed Source Coding, the aim is achieving the minimum data rate for coding an input source X less then a fidelity criterion D, given Y, a noisy observation of the source available at the decoder only with i.i.d. $\sim p(x,y)$. For simplicity, if the data hiding capability is disabled in Fig. 1, such that $M = \emptyset$ and $W = X$, hence the problem is turned out to be a Distributed Source Coding System. In addition to this setup, we embed a digital watermark M to the input source X with a distortion constraint between the input source and the watermarked embedded signal W such that $E[(X-W)^2] \leq D_1$. The watermarked embedded signal W is compressed to a data rate $R(W)$ such that it can be decoded with a fixed distortion $E[(W-\hat{W})^2] \leq D_2$, given that the encoder has an access to the original data X, and decoder has an access to the noisy observation Y. The proposed system can be viewed as a quantization of the input signal X in the sense of embedding a watermark M as a function of X (or known as context dependent watermarking). The watermarked signal W is compressed with a syndrome coding or coloring for the distributed source coding. In the decoder side, the received color indexes or syndromes are decoded with the help of the side information Y, and afterwards the embedded watermark $\hat{M}$ is estimated by using $\hat{W}$, the output of the syndromes decoding, and Y the side information available at the decoder.

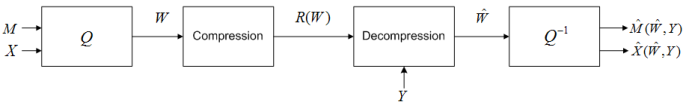

Figure 1: Data Hiding + Distributed Source Coding Scheme.

In Section.2, the duality between rate distortion and channel capacity with state information is given, both for a unique side information and the general case where there exists two different side information. A hybrid scheme is proposed for data hiding within a Distributed Source Coding setup in Section.3. Finally a preliminary simulation results of the proposed system are given in Section.4.

2. DUALITY

2.1 Distributed Source Coding($RDSI_{01}$)

Distributed Source Coding can be viewed as Rate Distortion with side information available at the decoder, which is shown schematically in Fig. 2. The notation in [15] is used such that subindex 01 in $RDSI_{01}$ indicates the availability of

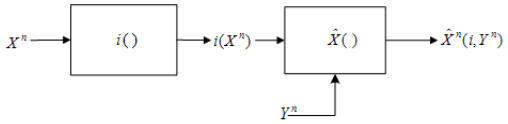

Figure 2: Rate distortion with side information available at the decoder.

a state information at the decoder but not the encoder. Let $\{(X_k,Y_k)\}$ i.i.d. $\sim p(x,y)$ be a sequence of independent drawings of jointly distributed random variables X and Y. X_k is encoded with block length n into a binary stream of rate R, by using a sequence of $(2^{nR},n)$ codes with $i:X^n \rightarrow \{1,2,..,2^{nR}\}$ and $\hat{X}^n : \{1,2,..,2^{nR}\} \rightarrow \hat{X}^n$. The input source X is to be encoded and transmitted to a receiver which access to a noisy observation Y, and $\hat{X}$ is estimated with a fidelity criterion D such that $E[(X,\hat{X})^2] \leq D$. The minimum rate of encoding [3] for a given fidelity criteria D is:

$$R_{01}(D) = \min_{\hat{X}=f(U;Y),p(u|x)} [I(U;X) - I(U;Y)] \qquad (1)$$

where the minimization is over all conditional probability density functions $p(u|x)$ and a function $f(U;Y)$ such that $E(X-\hat{X})^2 \leq D$. U is defined as an auxiliary variable for the set of codewords representing X and $I(U;X)$ is the mutual information between U and X.

2.2 Informed Data Hiding $(CCSI_{10})$

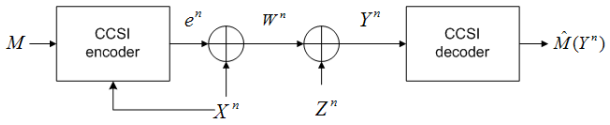

Figure 3: Channel coding with side information available at the encoder.

The blind watermarking problem can be viewed as channel coding with side information at the encoder which is shown in Fig 3. The encoder has access to a discrete watermark signal to be embedded M, and the host signal X that the information is to be embedded in. There is a fixed distortion constraint between the host signal X and the watermarked signal W such that $E(X-W)^2 \leq D_1$. Since $W = X + e$, and the error e is dependent on the input source X and M, where this setup is also known as content dependent data hiding. Then, the watermark embedded signal W is subjected to a fixed distortion attack Z. The achievable capacity [5] of the watermarking system for an error probability $P_e^n = Pr\{\hat{M}(Y^n,X^n) \neq M\}$ is:

$$C_{10} = \max_{p(u,w|x)} [I(U;Y) - I(U;X)] \qquad (2)$$

where U is an auxiliary variable and the maximization is over all conditional probability density function $p(u,w|x)$ and $I(U;Y)$ is the mutual information between U and Y. A rate R is achievable if there exists a sequence of $(2^{nR},n)$ codes with $P_e^n \rightarrow 0$. [15]

3. A GENERALIZED HYBRID SCHEME: DATA HIDING WITHIN DISTRIBUTED SOURCE CODING SYSTEM

In this section, we propose a hybrid scheme which utilize both channel coding and rate distortion with state information at the encoder and decoder respectively. The proposed system can be seen in Figure 4. Actually the system enables to hide the data M within a input host signal X with a distortion measure D_1. Then the watermarked embedded signal W is compressed and transmitted with a fidelity criterion D_2 to a receiver which has access to a noisy observation Y. Hence the encoder has access to two sources, the data to be hide or watermark index M and the input data source X that the information will be embedded in. The first criterion is embedding the watermark to the host signal X with a fixed-distortion measure D_1 where $E[(X-W)^2] \leq D_1$. Afterwards, for a minimum rate R, the embedded watermark signal W is coded and transmitted to a receiver, given that the encoder has access to the input source X, while decoder has an access to Y, which is a noisy observation of the input source X. We realize a JPEG attack channel between the input source X and the side informaton Y.The receiver decodes the received signal with the help of noisy observation Y with a fidelity criterion D_2 such that $E[(\hat{W}-W)^2] \leq D_2$ and estimates the watermarked signal $\hat{M}$ with an error probability $P_e(\hat{M})$

Mathematically, the goal is to solve the following constrained problem:

$$min_{E[(X-W)^2] \leq D_1, E[(\hat{W}-W)^2] \leq D_2} P_e(\hat{M}) \qquad (3)$$

where $P_e(\hat{M})$ represents the decoding error probability $Pr\{(\hat{M}(\hat{W},Y) \neq M)\}$, and W,X,Y are jointly distributed random variables with i.i.d. $p(w,x,y)$. Moreover the distortion constraint $E[(\hat{W}-W)^2] \leq D_2$ leads to a minimum rate function:

$$R(D_2) = \min_{p(u|w,x)p(\hat{w}|u,y)} [I(U;X,W) - I(U;Y)] \qquad (4)$$

The hybrid problem can be posed as a kind of semi-blind watermarking scheme. The receiver has not access to the input host signal X in order to extract $\hat{M}$ from the watermarked signal W, but Y, a noisy observation of the input source X.

4. EXPERIMENTAL SETUP AND RESULTS

Up to this point, our focus has been on the theoretical aspects of the mixture of two problems Channel Coding with Side Information and Source Coding with Side Information. In this section, we consider a real system that implements the data hiding and compression codes in joint manner. We discuss the details of the creation of codes in our system and give preliminary results.

4.1 Generation of the Side Information and Hidden Message

For our experiments, we used 512×512 gray-scale Lena image. Firstly the Lena image is uniformly quantized to the $4bits/sample$ rate and served as the host signal X which is available only at the encoder. Afterwards a virtual attack channel is defined between the Lena image and the side information available at the decoder Y. For this purpose, a JPEG

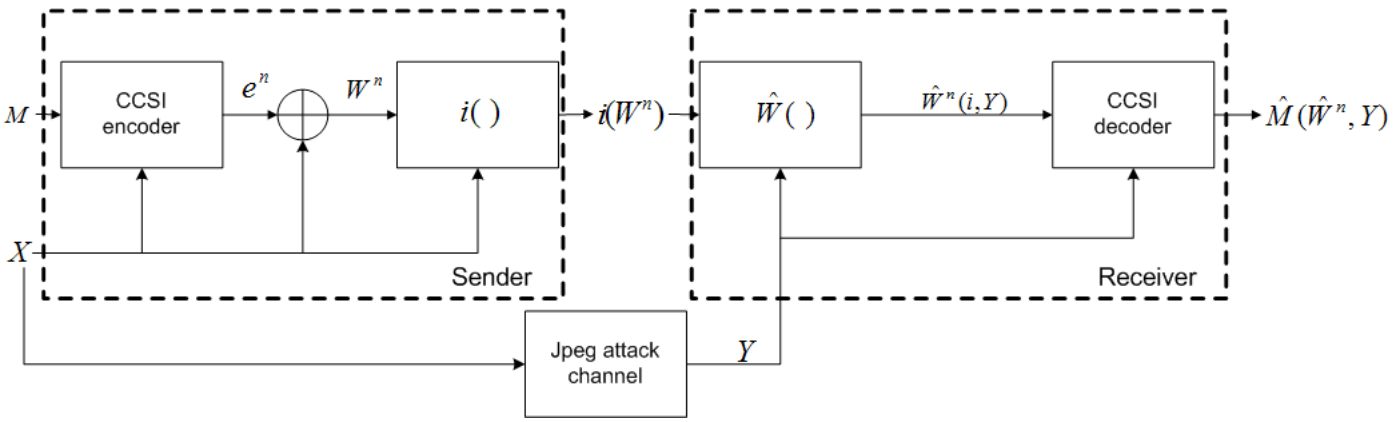

Figure 4: Proposed hybrid scheme: Data hiding within a Distributed Source Coding System.

compression attack is used to obtain Y from the original image. Several quality level of JPEG images are created started from 100% to 20%. The construction of the hidden message M to be embedded is i.i.d. pseudo-random Bernoulli$(1/2)$ string of appropriate block length so the first order-entropy of $H(M) = 1$bit/sample.

4.2 Data Hiding

For the case of informed data hiding of M within X, we used basic quantization based on memoryless coset construction. The algorithm is described as follows: 3 bits information is partitioned into 4 cosets such that each element of the coset has a hamming distance of 3. According to the two bits data of M the coset members of that index is chosen $Coset00 = \{000, 111\}$, $Coset01 = \{001, 110\}$, $Coset10 = \{010, 101\}$, $Coset11 = \{011, 100\}$. After creating the codebook, 2bits of M and R bits of X is taken. And the least significant 3 bits of the sub-block of the host signal X is depicted for embedding. The 3 bits value of X is quantized to $W : W(X,M) = \arg\min_{Z \in CosetM} \parallel Z - X \parallel$ which W is at most one bit differ from X. The distance metric is chosen as hamming distance. And this insertion of 2 bits within block length R continues until embedding all the data. As an example, assume that the 2 bits length message 01 is being embedded into the least 3 significant bits of X which is 010. The minimum hamming distance between 010 and the elements of $Coset01$ is chosen as the quantification output, which is $W = 110$ in this case. At the decoder side, the extraction of the watermark is straightforward such that the knowing the codebook and insertion frequency R, the coset index that the received block data resides in is decoded as the embedded data.

4.3 Compression and Decompression Setup

The watermarked string W is compressed by finding its syndrome with respect to a low-density parity check(LDPC) channel code[18]. In fact we use high-rate $(3/5)$ LDPC code and transmit only the check bits to the encoder. A code expressed as (n,k) where $m = n - k$ check bits are calculated from input stream of length k using a randomly generated parity check matrix H with dimension $(n-k) \times n$ whose codewords satisfy $Hx = 0$. Since only the parity check bits are sent to the decoder, so the compression rate is the number of check bits over input length $(n-k)/k$.

At the receiver, the compressed data is decoded using belief propagation algorithm in [7, 18], with the help of side information Y available only at the decoder. The goal of decoding is to find the nearest likelihood codeword $\hat{W}$ and extract the embedded string estimation $\hat{M}$. The side information is assumed to be the systematic bits and the received compressed data is assumed to be the parity checks. The belief propagation algorithm is very identical to that used for decoding standard LDPC codes, with some modifications to in our case. First, likelihood ratios of the systematic bits are initialized according to the correlation noise estimation between the side information. It is assumed to be that first order statistics of the two side information are available at the decoder. Second, initial likelihood ratios of the parity check bits are based on the fact that probability of received parity check is in error with a small probability ε. Moreover check-node update node is modified to recover the errors on the systematic bits using the fact that check-bits are correct with high probability. Finally, with the knowledge of the coset codebook and estimation of $\hat{W}$ using LDPC decoding its trivial to extract the hidden data $\hat{M}$. The distortion levels of the $\hat{W}$ and $\hat{M}$ are given in results.

4.4 Simulation Results

In our experiments, 4 bits/sample uniformly quantized Lena image is changed to 512 binary string blocks where each block is an vertical line of the image which has length 512×4. JPEG compressed side information with a quality constrained is also available at the decoder as in Sec.4.1. In this setup, a 1024 bits length hidden message is embedded as described in Sec.4.2 and watermarked signal W is obtained. while changing the correlation noise between side information. Afterwards W is compressed with 3/5 rate LDPC (3584,2048) code as explained in Sec.4.3. Since it is already quantized from 8 bits/sample to 4 bits/sample, after LDPC coding a total of 1/3 compression rate is achieved. Hence for each block, the 1536 parity check bits are sent to the receiver. And receiver decodes the received parity check bits using Y the side information available at the decoder as the systematic bits, using belief propagation with maximum of 50 iterations. Since the side information Y is seen to be the corrupted version of X, the parity check bits of W help to decode X correctly. The jpeg quality parameter is varied from

100% to 20% and plotted with their decoding error probabilities.

Figure 5: Simulation results.

As seen in Fig.5, for the case of our data hiding and distributed source coding scheme, the decompression of the coded signal is extracted perfecty for JPEG 80% channel which corresponds up to a 9% of bit errors between X and Y. According to the results on synthetic data in [1], up to 13% of bit errors can be corrected in Binary Symmetric Channel, however in JPEG attack channel we end up with 4% performance loss. Moreover, for each block the decoder system tries to find a most probable codeword, but our belief propagation algorithm does not always converge to a valid codeword within maximum iterations. In that case the decoded block has lower bit error probability but in overall the system output does not good in both PSNR and subjective tests.

Figure 6: Visualization of outputs, Left: decoding output for jpeg quality 50% SI. Right: The non-valid codewords of the decoding are replaced with SI blocks.

For instance you can see the output of decoding of lena image while side information is exposed to a quality of 50% in Fig.6 in left. When the decoder can not find a valid codeword for decoding, the performance of the system on that block decreases rapidly. The resulting image PSNR value of 3.47 dB. In our decoding strategy, if the result of the decoding block is an invalid codeword, we used the values of the Side Information that correspond to the falsely decoded block. The final decoded output can be seen in Fig.6 right which has a PSNR value of 21.30 dB. Please note that the PSNR value of the Side information is 19 dB.

5. CONCLUSIONS

In conclusion, we establish a hybrid system for hiding data to a compression process which uses distributed source coding system. Recent findings about the duality between rate distortion and channel capacity with state information are used for the system. The hybrid scheme proposed in Fig.4 offers a wide range of multi-source coding systems. The selection of the inputs such as the side information X,Y, and the nature of the hidden information M depends on the considered problem. We used a JPEG attack channel in order to construct the side information Y available at the decoder. A memoryless data hiding algorithm is used with LDPC based distributed compression scheme. A trellis based data hiding with memory can be used for improving the performance of the overall system. Indeed, this scheme can be easily adapted for example to video coding such that temporally correlated successive frames serve as sources S_i to be coded and the watermark signal is M to be embedded.

REFERENCES

[1] C. Dikici, K. Idrissi and A. Baskurt, “Joint Data Hiding and Source Coding with Partially Available Side Information”, *Proc. 18th SPIE Multimedia Processing and Applications Conference*, San Jose, USA, 2006.

[2] J. D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources”, *IEEE Transactions on Information Theory*, vol. IT-19, pp. 471–480, July 1973.

[3] A. D. Wyner and J. Ziv, “The Rate-Distortion Function for Source Coding with Side Information at the Decoder”, *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 110, Jan. 1976.

[4] M. Costa, “Writing on dirty paper,” IEEE Trans. on Information Theory, vol. 29, pp. 439–441, May 1983.

[5] S. Gel’fand and M. Pinsker, “Coding for channel with random parameters,” Problems of Control and Information Theory, vol. 9, pp. 19–31, 1980.

[6] C Berrou, “Turbo-codes, a breakthrough towards the Shannon limit”, Hughes Network Systems, Baltimore, Maryland, USA, July 1992.

[7] R. G. Gallager, Low density parity check codes, Ph.D. dissertation, MIT, Cambridge, MA, 1963.

[8] R. Puri and K. Ramchandran, “PRISM: A new robust video coding architecture based on distributed compression principles”, *Proc. Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, Oct. 2002.

[9] A. Aaron and B. Girod, “Compression with side information using turbo codes”, *Proc. IEEE Data Compression Conf.*, pp. 252–261, 2002.

[10] Cox I. J., M. L. Miller, and A. L. McKellips, Watermarking as communications with side information, Proceedings of the IEEE 87, pp. 11271141, July 1999.

[11] Eggers J., R. Buml, R. Tzschoppe and B. Girod, “Scalar costa scheme for information embedding”,IEEE Trans. Signal Processing,2002.

[12] Chappelier V., C. Guillemot and S. Marinkovic, “Turbo Trellis Coded Quantization,” Proc. of the Intl. symp. on turbo codes, September, 2003.

[13] J. Chou, S. S. Pradhan, and K. Ramchandran, “On the duality between distributed source coding and data hiding,” Proc. of 33rd Asilomar Conf. on Signals, Systems and Computers, November 1999.

[14] J. K. Su, J. J. Eggers and B. Girod, "Illustration of the Duality Between Channel Coding and Rate Distortion with Side Information," 34th Asilomar Conf. on Signals, Systems, and Computers. Oct. 29-Nov. 1, 2000, Asilomar, CA, USA.

[15] T. M. Cover and M. Chiang, 'Duality between channel capacity and rate distortion with two-sided state information,' IEEE Trans. of Inform. Theory, vol. 48, no. 6, pp. 1629 - 1638, June 2002.

[16] B. Rimoldi and R. Urbanke, "Asynchronous SlepianWolf coding via sourcesplitting," Proc. IEEE Int. Symp. on Info. Theory, Ulm, Germany, July 1997.

[17] T.J. Flynn and R.M. Gray, "Encoding of correlated observations", *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, Nov. 1987.

[18] MacKay, D. J. C. and R.M. Neal,"Near Shannon limit performance of low density parity check codes",*Electronics Letters,* vol. 33, pp. 457-458, 1996.