September 2002

Decoding of images using soft-bits and Markov random field modeling

Alfred Mertins
University of Wollongong, mertins@uow.edu.au

O. Jamart
University of Wollongong

Decoding of images using soft-bits and Markov random field modeling

Abstract

We combine soft-bit decoding with Markov random field modeling of source signals and apply the proposed technique to image communication over noisy channels. No redundancy is added by the encoder in the form of channel codes. For error correction and concealment, the decoder relies only on the bit-reliability information extracted at the channel output and the random field modeling of natural images. The results obtained show that the proposed method yields excellent performance even under extremely noisy conditions.

Disciplines

Physical Sciences and Mathematics

Publication Details

This was published as: Mertins, A & Jamart, O, Decoding of images using soft-bits and Markov random field modeling, Proceedings International Conference on Image Processing, 22-25 September 2002, 1, I-241- I-244.

DECODING OF IMAGES USING SOFT-BITS AND MARKOV RANDOM FIELD MODELING

Alfred Mertins, Olivier Jamart

University of Wollongong
School of Electrical, Computer, and Telecommunications Engineering
Wollongong, NSW 2522, Australia. Email: mertins@uow.edu.au.

ABSTRACT

In this paper, we combine soft-bit decoding with Markov random field modeling of source signals and apply the proposed technique to image communication over noisy channels. No redundancy is added by the encoder in form of channel codes. For error correction and concealment the decoder relies only on the bit-reliability information extracted at the channel output and the random field modeling of natural images. The results obtained show that the proposed method yields excellent performance even under extremely noisy conditions.

1. INTRODUCTION

Classical communication systems use separate units for source compression, error protection, and transmission over noisy channels. In recent years, however, combined source-channel coding techniques and soft-bit decoding algorithms have proven to yield better performance in many practical situations than the conventional designs [1–5]. In addition, error concealment methods, especially in video communications, have become part of most decoders for noisy environments or channels with erasures [6–8]. In this paper, we combine the soft-bit decoding technique introduced in [2] for speech decoding and applied in [3] to image decoding with Markov random field (MRF) modeling of the source to either correct bit errors or to minimize their effect on the reconstructed signal. MRF models are the basis for many error concealment and image restoration techniques [6–9], but thus far they have not been combined with soft-bit decoding.

The paper is organized as follows. In Section 2 the transmission model is introduced, and Section 3 presents the different decoding principles. Section 4 then describes the MRF model, and Section 5 presents results for the proposed decoding techniques. Finally, Section 6 gives some conclusions.

2. THE TRANSMISSION MODEL

Consider the transmission of the elements $X_{i,j}$, $i = 0,\ldots,M-1$, $j = 0,\ldots,N-1$, of an $M \times N$ matrix $\boldsymbol{X}$ which may, for example, represent an entire image or the quantized subband samples of the wavelet transform of an image. The symbols $X_{i,j}$ are assumed to be taken from an alphabet of 2^B possible symbols and can thus be represented with B bit code vectors $\boldsymbol{x}_{i,j} = [x_{i,j}(0), x_{i,j}(1), \ldots, x_{i,j}(B-1)]^T$ where $x_{i,j}(m)$, $m = 0, 1, \ldots, B-1$ are the actual bits representing $X_{i,j}$. In view of a BPSK transmission of $x_{i,j}(m)$, we assume bipolar bits, i.e. $x_{i,j}(m) \in \{-1, +1\}$.

The bits $x_{i,j}(m)$ are assumed to be sequentially transmitted across an additive white Gaussian noise (AWGN) channel with two-sided noise power spectral density $N_0/2$. The received samples, denoted as $y_{i,j}(m)$, can then be written as

$$y_{i,j}(m) = x_{i,j}(m) + n_{i,j}(m) \tag{1}$$

where $n_{i,j}(m)$ are zero-mean, statistically independent Gaussian random variables with variance σ_n^2. Because of $x_{i,j}(m) \in \{-1, +1\}$ the energy used to transmit one bit (E_b) is one, and the often used E_b/N_0 ratio becomes $E_b/N_0 = 1/(2\sigma_n^2)$.

3. DECODING PRINCIPLES

In this section we will look at three decoding principles: hard decoding, maximum a posteriori (MAP) soft decoding, and non-linear, mean squares (MS) estimation based on soft bit information.

A hard decoder will simply carry out a threshold detection of the individual received bits using the rule

$$\hat{y}_{i,j}(m) = \begin{cases} +1 & \text{if } y_{i,j}(m) \geq 0 \\ -1 & \text{if } y_{i,j}(m) < 0 \end{cases} \tag{2}$$

and will set the final bit-wise decisions $\hat{x}_{i,j}(m)$ to $\hat{x}_{i,j}(m) := \hat{y}_{i,j}(m)$. The codeword $\hat{\boldsymbol{x}}_{i,j} = [\hat{x}_{i,j}(0), \ldots, \ldots, \hat{x}_{i,j}(B-1)]^T$ then corresponds to an estimate $\hat{X}_{i,j}$ for the transmitted symbol $X_{i,j}$. Hard decoding is usually accompanied with error control coding, because otherwise, a few bit errors can lead to a significant performance degradation.

The concept of soft bits is to forward the probabilities with which the various hard decisions are correct or incorrect to the decoding algorithm, instead of making final hard decisions straight away. The required probabilities can be computed from the instantaneous bit error rate of hard decoding given by [2]

$$P_e(i,j,m) = \left(1 + e^{4|y_{i,j}(m)|E_b/N_0}\right)^{-1}. \tag{3}$$

As one might expect, the instantaneous bit error rate is low when the received value $|y_{i,j}(m)|$ is large, and it tends to 0.5 for $|y_{i,j}(m)| \to 0$.

Maximum a posteriori decoding:

A MAP detector decides for the code vector (or symbol) with the highest *a posteriori* probability among all 2^B possible code vectors (symbols). With

$\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{2^B-1}$ denoting the 2^B hypotheses and $\hat{\boldsymbol{y}}_{i,j} = [\hat{y}_{i,j}(0), \hat{y}_{i,j}(1), \ldots, \hat{y}_{i,j}(B-1)]^T$ being the vector of hard decoded bits the decision rule can be written as

$$\hat{x}_{i,j} = \tilde{x}_K \quad \text{with} \quad K = \arg \max_{k=0}^{2^B-1} P(\tilde{\boldsymbol{x}}_k | \hat{\boldsymbol{y}}_{i,j}). \tag{4}$$

Using the Bayes rule the required *a posteriori* probabilities can be expressed as

$$P(\tilde{\boldsymbol{x}}_k | \hat{\boldsymbol{y}}_{i,j}) = C \cdot P(\hat{\boldsymbol{y}}_{i,j} | \tilde{\boldsymbol{x}}_k)\, P(\tilde{\boldsymbol{x}}_k) \tag{5}$$

where $P(\tilde{\boldsymbol{x}}_k)$ is the *a priori* probability of vector $\tilde{\boldsymbol{x}}_k$, and $P(\hat{\boldsymbol{y}}_{i,j} | \tilde{\boldsymbol{x}}_k)$ is the transition probability for receiving $\hat{\boldsymbol{y}}_{i,j}$ after $\tilde{\boldsymbol{x}}_k$ has been sent. C is a normalization constant. The transition probabilities $P(\hat{\boldsymbol{y}}_{i,j} | \tilde{\boldsymbol{x}}_k)$ can be written as a product of the bit-wise transition probabilities $P(\hat{y}_{i,j}(m) | \tilde{x}_k(m))$ as

$$P(\hat{\boldsymbol{y}}_{i,j} | \tilde{\boldsymbol{x}}_k) = \prod_{m=0}^{B-1} P(\hat{y}_{i,j}(m) | \tilde{x}_k(m)) \tag{6}$$

where the terms $P(\hat{y}_{i,j}(m) | \tilde{x}_k(m))$ are found from the instantaneous bit error rate:

$$P(\hat{y}_{i,j}(m) | \tilde{x}_k(m)) = \begin{cases} 1 - P_e(i,j,m) & \text{if } \tilde{x}_k(m) = \hat{y}_{i,j}(m), \\ P_e(i,j,m) & \text{if } \tilde{x}_k(m) \neq \hat{y}_{i,j}(m). \end{cases} \tag{7}$$

The advantage of the rule (4) with the probabilities according to (5) over hard decoding becomes obvious when the largest transition probabilities for two ore more hypotheses are in the same range. Then the hypothesis with the highest *a priori* probability will be favored by the decoder, because this will most likely be the correct one.

Equation (4) only contains zero-order *a priori* knowledge about the code vectors. However, the *a priori* knowledge can be easily extended to higher order by taking the neighboring symbols into account. We will look into this in the next section.

Mean-squares estimation:

Given the *a posteriori* probabilities for the various hypotheses the MS symbol estimate can be found as

$$\hat{X}_{i,j} = \sum_{k=0}^{2^B-1} \tilde{X}_k\, P(\tilde{\boldsymbol{x}}_k | \hat{\boldsymbol{y}}_{i,j}) \tag{8}$$

where $\tilde{X}_k$ is the symbol corresponding to code vector $\tilde{\boldsymbol{x}}_k$. This type of estimation is useful when the symbols $X_{i,j}$ are numerical values, because then $\hat{X}_{i,j}$ is the estimated numerical value with minimum mean-squared error. Minimizing this error is then equivalent to maximizing the final signal-to-noise ratio.

4. MARKOV RANDOM FIELD MODELING OF A PRIORI KNOWLEDGE

Fingscheidt and Vary [2] considered the transmission of one-dimensional signals (speech). The transmitted symbols were modeled as 0th order and 1st order Markov processes, respectively. This concept was extended to two dimensions by Kliewer and Görtz in [3]. They used a neighborhood of four symbols (i.e. $X_{i,j-1}, X_{i,j+1}, X_{i-1,j}, X_{i+1,j}$) and estimated the transition probabilities from these neighboring symbols to the actual symbol $X_{i,j}$ either from the image itself or from a large set of representative images (only the latter is suitable in practice). This *a priori* knowledge was then stored and used during decoding.

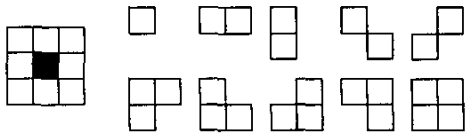

Fig. 1. Eight-pixel neighborhood system and all associated cliques.

In this paper, we model the *a priori* knowledge about the symbols via Markov random fields (MRF) with a neighborhood system of eight neighbors. An advantage of this approach over the one in [3] is that no transition probabilities need to be stored or transmitted. For details on Markov random field theory the reader is referred to [9]. Here will only describe their use in a straight forward manner.

Due to the well-known Markov-Gibbs equivalence a probability $P(x)$ for an element of a Markov random field can be described as a Gibbs distribution [9]:

$$P(x) = Z^{-1}\, e^{-\frac{1}{T}U(x)} \tag{9}$$

where $U(x)$ is called an energy function, T is called the temperature, and Z is known as the partition function. The energy function can be written as a sum over potential functions $V_C(x)$ for all cliques belonging to the neighborhood of x:

$$U(x) = \sum\nolimits_C V_C(x). \tag{10}$$

An example for a neighborhood system and the associated cliques is shown in Fig. 1. The partition function Z is used for normalization. It is given by

$$Z = \sum\nolimits_x e^{-\frac{1}{T}U(x)} \tag{11}$$

and ensures that the probabilities for all possible choices for x sum up to one.

The choice of the potential function is crucial, because it describes the properties of the Markov random field. Many potential functions suitable for different types of images and applications have been proposed in the literature. Our experiments have shown that the function

$$V_C(X_{i,j}, X_{k,\ell}) = \log(\delta^2 + (X_{i,j} - X_{k,\ell})^2) + \frac{1}{\delta^2 + (X_{i,j} - X_{k,\ell})^2} - \log \delta^2 - \frac{1}{\delta^2}, \tag{12}$$

which has been proposed in [7] for decoding of noisy images, together with the two-site cliques in Fig. 1 is a good choice for soft decoding of images. The choices for the parameters δ and

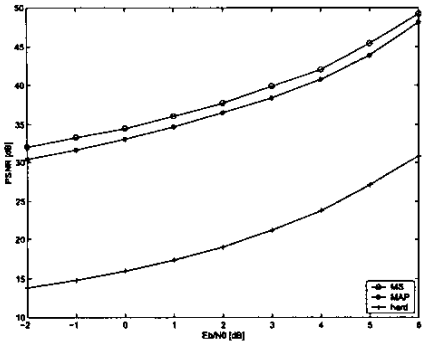

Fig. 2. PSNR results for transmission of the original image and different decoding methods.

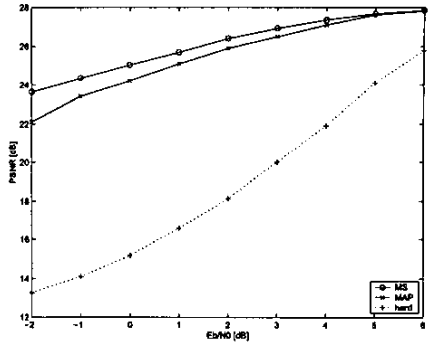

Fig. 3. PSNR results for transmission of wavelet coefficients at 0.36 bpp and different decoding methods.

T will be presented in Section 5 together with the discussion of experimental results.

Like in classical MRF-based image restoration (see e.g. [9] we use an iterative decoding approach. That is, the reconstruction algorithm is applied multiple times, until convergence is achieved and the reconstructed values do not change anymore. However, contrary to classical image restoration we do not recompute the transition probabilities from the original pixels to the observed ones in each iteration step. Instead, we use the transition probabilities $P(\hat{y}_{i,j}(m)|\tilde{x}_k(m))$ derived from the soft-bit information according to (7).

5. EXPERIMENTAL RESULTS

We consider two experimental settings. In the first one, we transmit the original pixels of an image in PCM over the channel. In a second setting we consider the transmission of the coefficients of the wavelet transform of the same image. The image chosen is the Goldhill image of size 512×512 pixels with 8 bits per pixel (bpp).

For modeling the *a priori* density of the original image the parameters T and δ were chosen as $T = 1.9$ and $\delta^2 = 50$. No temperature annealing, as often used in image restoration, was carried out. The E_b/N_0 ratio was assumed to be known to the receiver. To speed up the recursive decoding procedure, only sites for which the maximal *a posteriori* probability was below 0.95 were revisited during the next iterations. For an E_b/N_0 of 4 dB and above this meant that 75 % of the sites were visited only once. In our experiment the E_b/N_0 ratio was varied between -2 and 6 dB, and the peak signal-to-noise ratio (PSNR) was recorded. Fig. 2 shows the results for hard, MAP, and MS decoding. As one can see, the MS decoder yields the best results, whereas hard decoding performs extremely poor.

In a second set of experiments the image was first transformed with a three-level wavelet transform based on the 9-7 wavelet of [10]. Symmetric reflection was used at the image boundaries. The wavelet coefficients were quantized such that the total transmitted bit rate amounts to 0.36 bpp including all side information. The bit allocation as well as the wavelet transform and the image were the same as in [3], to enable comparisons between our MRF-based decoder and the one in [3]. The MRF parameters were chosen as $T = 2.8$, $\delta = 1$ for the lowest frequency (LL) band and as $T = 4$, $\delta = 1$ for all other bands. Each band was decoded separately. To speed up computation significantly, sites with an MAP probability above 0.95 for the LL band and above 0.98 for the other bands were not revisited.

Fig. 3 shows the PSNR's for the various reconstruction methods. The maximum PSNR for error-free transmission is 27.9 dB. Again, the MS estimation performs best, and the maximum possible PSNR is almost reached at an E_b/N_0 ratio of 6 dB. A comparison with the results in [3] for a realistic scenario where the transition probabilities were obtained from a training set shows that our MRF-based decoders are superior for almost all E_b/N_0 ratios. At an E_b/N_0 of -2 dB our MAP decoder is about 1.25 dB better than the one in [3] and the MS decoder is even superior by 2 dB. For an E_b/N_0 of 6 dB all decoders almost reach the maximum achievable PSNR. Interestingly, at an E_b/N_0 of -2 dB our MS decoder is even better than the one in [3] for the case where the probability model was derived from the transmitted image itself. A reason for the good performance of the MRF decoder seems to be the fact that the decoding is carried out iteratively, changing the decoded symbols multiple times until the most likely combination of decoded symbols is found.

To get a visual impression of the performance of the proposed soft-decoding algorithms, Figs. 4 and 5 show examples of hard and soft decoded images at an E_b/N_0 of 0 dB.

6. CONCLUSIONS

We have introduced a robust image decoding method that combines MRF modeling of *a priori* information about natural images with bit-reliability information extracted at the channel output. Even on channels with extremely low E_b/N_0 ratio the decoder performs very well. In particular, it significantly outperforms the technique in [3] under extremely noisy conditions. While the experiments presented here have been carried

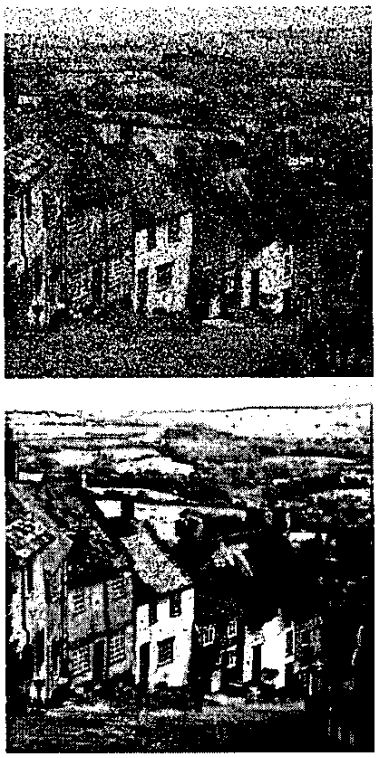

Fig. 4. Examples of decoded images where the original 8 bpp image was transmitted over a channel with an E_b/N_0 of 0 dB. Top: hard decoding; bottom: MS soft decoding.

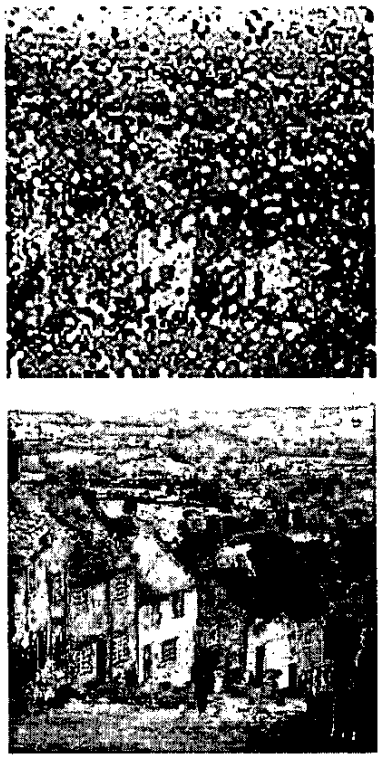

Fig. 5. Examples of decoded images where the DWT coefficients of the image were transmitted at a rate of 0.36 bpp over a channel with an E_b/N_0 of 0 dB. Top: hard decoding; bottom: MS soft decoding.

out for subband-wise fixed-length codes, future work will be directed toward decoding of variable length codes with MRF source modeling.

7. REFERENCES

[1] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sept. 1995.

[2] T. Fingscheidt and P. Vary, "Robust speech decoding: A universal approach to bit error concealment," in *Proc. ICASSP*, Munich, Germany, Apr. 1997, pp. 1667–1670.

[3] J. Kliewer and N. Görtz, "Soft-input source decoding for robust transmission of compressed images using two-dimensional optimal estimation," in *Proc. ICASSP*, Salt Lake City, Utah, USA, May 2001, vol. 4, pp. 2565–2568.

[4] M. Skoglund and P. Hedelin, "Hadamard-based soft decoding for vector quantization over noisy channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 515–532, Mar. 1999.

[5] N. Phamdo and F. Alajaji, "Soft-decision demodulation design for COVQ over white, colored and ISI Gaussian channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1499–1506, Sept. 2000.

[6] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[7] O. K. Al-Shaykh and R. M. Mersereau, "Lossy compression of noisy images," *IEEE Trans. Image Processing*, vol. 7, no. 12, pp. 1641–1652, Dec. 1998.

[8] S. Shirani and F. Kossentini, "A concealment method for video communications in an error-prone environment," *IEEE Journal on Selected Areas in Comm.*, vol. 18, no. 6, pp. 1122–1128, June 2000.

[9] S. Geman and D. Geman, "Stochastic relation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–740, Nov. 1984.

[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992.