Image Recovery from Noisy Transmission using Soft Bits and Markov Random Field Models

Alfred Mertins

University of Oldenburg

School of Mathematics and Natural Sciences

D-26111 Oldenburg, GERMANY

Telephone +49 441 798 3610; Facsimile +49 441 798 3902

Email: alfred.mertins@uni-oldenburg.de

Abstract

In this paper, a new method for robust image recovery from noisy transmission based on soft-bits and Markov random field modeling is proposed. The method aims to exploit the residual redundancy present in the symbols produced by a source encoder together with knowledge about the statistical properties of natural images. The soft-bit information required for reconstruction may either be extracted directly at the output of a noisy communication channel or at the output of a soft channel decoder (if channel codes are used). In the latter case the proposed algorithm may also be used in conjunction with iterative channel decoding, resulting in joint source-channel turbo decoding. Examples will be presented for the case where no channel codes are employed. The results obtained indicate that the proposed method yields very good performance even under extremely noisy conditions.

PACS numbers:

Keywords: Soft decoding, Markov random fields, image transmission, bit reliability, error concealment, source-channel coding

I. INTRODUCTION

Classical communication systems use separate units for source compression, error protection, transmission over noisy channels, and error concealment. 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 [1–12].

An often used approach is unequal error protection where more important information is protected through stronger channel codes, typically realized in form of rate compatible punctured convolutional codes [5, 12, 13]. Given a fixed amount of added redundancy, the influence of channel errors on the perceptual quality can then be minimized through a proper assignment of redundancy.

In addition to channel codes, post processing error concealment methods have become part of most decoders for noisy environments and channels with erasures [14–18]. This is especially the case in video communications, where information from previous frames together with motion vectors is used to conceal erroneous macroblocks in current frames. Techniques applicable to block-wise transmission of single images estimate the content of missing or damaged blocks from adjacent blocks using smoothness constraints or Bayesian approaches [16–18]. Bayesian image restoration was introduced by Geman and Geman in [19], who applied Markov random field theory and the Markov-Gibbs correspondence for describing the *a priori* information on images in order to obtain maximum *a posteriori* (MAP) image estimates. Their method and many of the follow-up methods are suitable for general image restoration applications and not only for error concealment in image/video transmission.

Fingscheidt and Vary used soft bits and residual codeword redundancies for error concealment of speech signals [2]. Based on 0th or 1st order Markov models for the transmitted symbols they derived mean squares (MS) and maximum a posteriori (MAP) estimators for the speech waveform. Their methods were quite effective, although they did not include channel codes. The concept of [2] was extended by Kliewer and Görtz in [4] to two dimensions. They collected data about the transition probabilities from the neighboring symbols (pixels) to the actual symbol and used this information as *a priori* information during the decoding process. The data was collected either from the transmitted image itself or from a large set of representative images. Only the latter strategy proved to be suitable in practice, because otherwise the transition probabilities would have to be transmitted as side information. Another method based on hidden Markov mesh models was proposed in [9]. Other approaches for error concealment using soft-bit information consider vector quantization of the source [6–8]. In [5] Kliewer and Görtz extended their work and combined the soft-bit based concealment technique with channel coding, resulting in a decoder structure that is similar to structures for decoding concatenated channel codes.

In this paper, we combine the soft-bit decoding technique of [4, 5] with source modeling based on Markov random fields (MRF's). The decoding procedure is iterative and mainly similar to the one used in image restoration [19]. However, other than in classical image restoration, we obtain the required transition probabilities directly from the available soft-bits and not from the observed image pixels. The iteration affects only the estimated *a priori* knowledge. Similar to [5], the proposed decoder can be combined with channel codes, resulting in joint source-channel turbo decoding. To demonstrate the decoder properties, experimental results are presented for the case without channel codes, where the entire concealment process relies on residual symbol redundancies and proper source modeling.

The paper is organized as follows. In Section II the transmission model is introduced, and Section III presents the different decoding principles. Section IV then describes the MRF model, and Section V presents results for the proposed decoding techniques. Finally, Section VI gives some conclusions.

II. THE TRANSMISSION MODEL

We consider a general transmission model in which the elements $X_{i,j}$, $i = 0, \ldots, M-1$, $j = 0, \ldots, N-1$ of an $M \times N$ matrix $\boldsymbol{X}$ are transmitted through a noisy channel. The elements $X_{i,j}$ may, for example, represent the quantized pixels of an image or the quantized subband samples of the wavelet transform of an image. More general, they may also be the indices describing codebook entries of a vector quantizer.

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 bits. With $x_{i,j}(m)$, $m = 0, 1, \ldots, B-1$ being the actual bits representing $X_{i,j}$ we can form the code vectors

$$\boldsymbol{x}_{i,j} = [x_{i,j}(0), x_{i,j}(1), \ldots, x_{i,j}(B-1)]^T.$$

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 considered 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 E_b used to transmit one bit is equal to one, and the often used E_b/N_0 ratio becomes $E_b/N_0 = 1/(2\sigma_n^2)$.

III. DECODING PRINCIPLES

In this section we will look at three different decoding principles. These are hard decoding, MAP soft decoding, and non-linear MS estimation based on soft bit information. Hard decoding is mainly used as a reference.

A. Hard decoding

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

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

The final bit-wise decisions $\hat{x}_{i,j}(m)$ will simply be set to $\hat{x}_{i,j}(m) := \hat{y}_{i,j}(m)$. The codeword

$$\hat{\boldsymbol{x}}_{i,j} = [\hat{x}_{i,j}(0), \hat{x}_{i,j}(1), \ldots\ldots, \hat{x}_{i,j}(B-1)]^T$$

then corresponds to an estimate $\hat{X}_{i,j}$ for the transmitted symbol $X_{i,j}$.

B. Soft bits and reliability

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 [1]

$$P_e(i,j,m) = \frac{1}{1+\exp(L(m)\cdot y_{i,j}(m))} \tag{3}$$

with $L(m)$ being the log-likelihood ratio

$$L(m) = \ln \frac{P(x_{i,j}(m) = +1|y_{i,j}(m))}{P(x_{i,j}(m) = -1|y_{i,j}(m))}. \tag{4}$$

For the case of BPSK in Gaussian noise considered in this paper, the log-likelihood ratio is given by $L(m) = 4E_b/N_0$, so that the instantaneous bit error rate becomes

$$P_e(i,j,m) = \frac{1}{1+\exp(4|y_{i,j}(m)|E_b/N_0)}. \tag{5}$$

According to (5), the instantaneous bit error rate is low when $|y_{i,j}(m)|$ (i.e. the magnitude of the received value) is large, and it tends to 0.5 for $|y_{i,j}(m)| \to 0$. The expected value of $P_e(i,j,m)$ is the bit error rate of BPSK as e.g. derived in [20]. The values $P_e(i,j,m)$ can be seen as bit reliability information that is attached to the hard decisions $\hat{y}_{i,j}(m)$ and can be used by more sophisticated decoders to obtain better quality than that provided through hard decoding.

C. Maximum a posteriori decoding

Among all 2^B possible code vectors (symbols) a MAP detector decides for the one with the largest *a posteriori* probability. With $\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{2^B-1}$ denoting the 2^B hypotheses (i.e. the vectors of possible bit combinations) 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

$$\hat{\boldsymbol{x}}_{i,j} = \tilde{\boldsymbol{x}}_K \quad \text{with} \quad K = \arg\max_{k=0}^{2^B-1} P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j}). \tag{6}$$

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

$$P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j}) = C \cdot P(\hat{\boldsymbol{y}}_{i,j}|\tilde{\boldsymbol{x}}_k)\, P(\tilde{\boldsymbol{x}}_k) \tag{7}$$

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 that ensures that the integral over $P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j})$ is one. Assuming statistically independent data bits and noise samples, the transition probabilities $P(\hat{\boldsymbol{y}}_{i,j}|\tilde{\boldsymbol{x}}_k)$ can be written as products of the bit-wise transition probabilities $P(\hat{y}_{i,j}(m)|\tilde{x}_k(m))$ for $m = 0, 1, \ldots, B-1$. This yields

$$P(\hat{\boldsymbol{y}}_{i,j}|\tilde{\boldsymbol{x}}_k) = \prod_{m=0}^{B-1} P(\hat{y}_{i,j}(m)|\tilde{x}_k(m)). \tag{8}$$

The transition probabilities $P(\hat{y}_{i,j}(m)|\tilde{x}_k(m))$ are found from the instantaneous bit error rate as follows:

$$P(\hat{y}_{i,j}(m)|\tilde{x}_k(m)) = \begin{cases} 1 - P_e(i,j,m) & \text{if } \tilde{x}_k(m) = \hat{y}_{i,j}(m), \\ P_e(i,j,m) & \text{if } \tilde{x}_k(m) \neq \hat{y}_{i,j}(m). \end{cases} \tag{9}$$

The advantage of the rule (6) with the probabilities according to (7) 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.

D. Mean-squares estimation

Given the *a posteriori* probabilities for the various hypotheses $\tilde{\boldsymbol{x}}_k$ the MS symbol estimate can be found as [21]

$$\hat{X}_{i,j} = \sum_{k=0}^{2^B-1} \tilde{X}_k\, P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j}) \tag{10}$$

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

E. Higher order a priori knowledge

Equations (6) and (10) contain only 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. Then $P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j})$ in (6) or (10) needs to be replaced by $P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j}, N_{\hat{\boldsymbol{y}}_{i,j}})$ where $N_{\hat{\boldsymbol{y}}_{i,j}}$ is the neighborhood of $\hat{\boldsymbol{y}}_{i,j}$. The Bayes rule then formally yields

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

Because for the AWGN channel $P(\hat{\boldsymbol{y}}_{i,j}|\tilde{\boldsymbol{x}}_k, N_{\hat{\boldsymbol{y}}_{i,j}})$ is independent of $N_{\hat{\boldsymbol{y}}_{i,j}}$ we finally obtain the expression

$$P(\tilde{\boldsymbol{x}}_k|\hat{\boldsymbol{y}}_{i,j}, N_{\hat{\boldsymbol{y}}_{i,j}}) = C \cdot P(\hat{\boldsymbol{y}}_{i,j}|\tilde{\boldsymbol{x}}_k)\, P(\tilde{\boldsymbol{x}}_k|N_{\hat{\boldsymbol{y}}_{i,j}}) \qquad (11)$$

that replaces (7). Methods for describing $P(\tilde{\boldsymbol{x}}_k|N_{\hat{\boldsymbol{y}}_{i,j}})$ will be discussed in the next section.

IV. MARKOV RANDOM FIELD MODELING OF A PRIORI KNOWLEDGE

In this paper, we model the *a priori* knowledge about the symbols via Markov random fields using the well-known Markov-Gibbs correspondence. For this we consider a neighborhood system of the eight nearest neighbors. An advantage of this approach over the one in [4] is that no transition probabilities need to be stored or transmitted. The same random filed model can be used for all images. For details on Markov random field theory the reader is referred to the work by Geman and Geman [19] or the books [22, 23] on MRF's. In the following we will describe their use in a straight forward manner.

Expressing the statistical properties of MRF's directly in form of conditional probabilities can be quite difficult. Fortunately, due to the Markov-Gibbs equivalence established in the Hemmersley-Clifford theorem [24], any MRF can be completely described via a Gibbs distribution. This means that a probability $P(x)$ for an element of a Markov random field can be written as

$$P(x) = Z^{-1}\, e^{-\frac{1}{T}U(x)}. \qquad (12)$$

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

$$U(x) = \sum\nolimits_C V_C(x). \tag{13}$$

Various neighborhood systems N_x of a site x can be defined. However, a neighborhood system must obey the following required properties:

(i) $x \notin N_x$; this means that x does not belong to its own neighborhood;

(ii) $x_i \in N_{x_j} \leftrightarrow x_j \in N_{x_i}$; this means that if x_i is in the neighborhood of x_j then x_j is in the neighborhood of x_i, and vice versa.

An example for a neighborhood system of eight neighbors and the associated cliques is shown in Fig. 1. The first clique consists of single sites. The second type of clique consists of x and its horizontal neighbors. The third type includes x and its vertical neighbors, and so on.

The partition function Z is used for normalization. For a discrete random variable x it is given by

$$Z = \sum\nolimits_x e^{-\frac{1}{T}U(x)}. \tag{14}$$

The multiplication with Z^{-1} in (12) simply 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. An overview of some of the most commonly used ones for pair-wise cliques is given in Table I.

To see the differences between the various potential functions we consider a simple example where a pixel x lies in the range $0, \ldots, 127$, having four neighbors given by 35, 45, 65, 100. The *a priori* probabilities $P(x|N_x)$ produced by the different potential functions are depicted in Fig. 2. The parameters for these potential functions were chosen such that the MS reconstruction of natural images performs best. As one can see, the Gaussian approach favors values x that are very close to the mean value of the neighbors. This consequently results in a very smooth reconstruction that lacks of detail and sharp edges. All other potential functions favor values in the range where most of the neighbors are, but not necessarily the arithmetic mean of the neighbors. Numerous experiments have shown that these functions perform much better on edges than the Gaussian one and that they also give good performance for both reconstructing images themselves as well as their wavelet transforms.

Like in classical Bayesian, MRF-based image restoration [19] we use an iterative decoding approach. In other words, 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 re-compute 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 (9).

V. EXPERIMENTAL RESULTS

In the following 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 grayscale image of size 512×512 pixels with 8 bits per pixel (bpp).

Experiments have been carried out for both of the above mentioned scenarios with all potential functions of Table 1, using properly chosen parameters p, δ, T. The best results were obtained with the generalized Gaussian approach according to Bouman and Sauer [25]. However, the peak signal-to-noise ratios (PSNR's) obtained with the other functions (except the Gaussian one) were not more than 0.5 dB below the one for the generalized Gaussian potential function. Because of the good performance and simplicity of the generalized Gaussian approach, further experimental results will be presented for this particular potential function. The normal Gaussian function either oversmoothed the reconstruction for low T or left significant artifacts for higher T. At low T it also occasionally led to numerical problems, because the values $e^{-U(x)/T}$ could become extremely small. The PSNR's obtained with the Gaussian function were several dB below the other ones.

For modeling the *a priori* density of the original image with the generalized Gaussian potential function the parameters T and p were chosen as $T = 4$ and $p = 0.7$, and only the four nearest neighbors were used. The temperature was kept constant during the iteration. 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 largest *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.

To demonstrate the convergence of the iterative decoding approach, the PSNR for the reconstructed image was recorded during the iteration. PCM transmission of the image across a channel with $E_b/N_0 = 0$ dB was considered. The results for the MS and MAP estimators are depicted in Fig. 3. The starting point for $n = 0$ corresponds to hard decoding. One can see that already a single iteration yields a significant improvement, and only a few iterations are required for convergence. The figure also shows the advantage of the MS over the MAP estimator.

In a second experiment for PCM transmission, the E_b/N_0 ratio was varied between -2 and 6 dB. Fig. 4 shows the obtained PSNR's for hard, MAP, and MS decoding. As one can see, the performance of soft decoding is very good, even for very noisy channels. Again, the MS decoder yields the best results, whereas hard decoding performs extremely poor.

To demonstrate the performance of the proposed soft decoding method for a scenario where less residual redundancy among the transmitted bits is present, the image was first transformed with a three-level wavelet transform based on the 9-7 wavelet of [26] and then quantized. 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 on image size, quantization step sizes, and wavelet tree depth. The bit allocation as well as the wavelet transform and the image were the same as in [4], to enable comparisons between our MRF-based decoder and the one in [4]. The side information was assumed to arrive error free at the decoder. The MRF parameters for decoding were chosen as $T = 5$ and $p = 0.7$. For the lowest frequency (LL) band only the four nearest neighbors were considered, whereas for all other bands a neighborhood of eight neighbors was used. Each band was decoded separately. To speed up the computation significantly, sites for which the largest *a posteriori* probability was above 0.95 for the LL band and above 0.98 for the other bands were not revisited. The slight increase in temperature compared to direct PCM transmission takes care of the fact that the wavelet coefficients have less statistical bindings than the original pixels of the image.

Fig. 5 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 [4] for a realistic scenario where the transition probabilities were obtained from a training set shows that the proposed MRF-based decoders are superior for small E_b/N_0 ratios and equivalent for higher ones. At an E_b/N_0 of -2 dB our MAP decoder is about 1.5 dB better than the one in [4] and the MS decoder is even superior by 2 dB. For an E_b/N_0 of 6 dB all decoders except the hard decoder 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 [4] 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. 6 and 7 show examples of hard and soft decoded images at an E_b/N_0 of 0 dB.

VI. 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 [4] under extremely noisy conditions. The proposed decoding method can be combined with channel codes, and decoding can then be carried out in the same way as the decoding of concatenated channel codes similar to [5]. While the experiments presented here have been carried out without channel coding and for subband-wise fixed-length codes, future work on MRF-based soft decoding will be directed toward decoding of variable length codes in combination with channel codes.

---

[1] J. Hagenauer, IEEE Trans. Commun. **43**, 2449 (1995).

[2] T. Fingscheidt and P. Vary, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Munich, Germany, 1997), pp. 1667–1670.

[3] T. Fingscheidt and P. Vary, IEEE Trans. Speech and Audio Processing **9**, 240 (2001).

[4] J. Kliewer and N. Görtz, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Salt Lake City, Utah, USA, 2001), vol. 4, pp. 2565–2568.

[5] J. Kliewer and N. Görtz, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Orlando, Fl, USA, 2002).

[6] M. Skoglund and P. Hedelin, IEEE Trans. Inform. Theory **45**, 515 (1999).

[7] N. Phamdo and F. Alajaji, IEEE Trans. Commun. **48**, 1499 (2000).

[8] M. Skoglund, IEEE Trans. Commun. **48**, 1309 (2000).

[9] M. Park and D. J. Miller, IEEE Trans. Image Processing **8**, 863 (1999).

[10] H. H. Otu and K. Sayood, IEEE Trans. Commun. **47**, 1615 (1999).

[11] N. Görtz, IEEE J. Sel. Areas in Commun. **14**, 1662 (2001).

[12] K. N. N. C. W. Yap and R. Liyanapathirana, Signal Processing: Image Communication **14**, 559 (1999).

[13] J. Hagenauer, IEEE Trans. Commun. **36**, 389 (1988).

[14] Y. Wang and Q. Zhu, Proceedings of the IEEE **86**, 974 (1998).

[15] O. K. Al-Shaykh and R. M. Mersereau, IEEE Trans. Image Processing **7**, 1641 (1998).

[16] S. Shirani, F. Kossentini, and R. Ward, IEEE Journal on Selected Areas in Comm. **18**, 1122 (2000).

[17] S. S. Hemami and T. H.-Y. Meng, IEEE Trans. Image Processing **4**, 1023 (1995).

[18] P. Salama, N. Shroff, and E. Delp, IEEE J. Sel. Areas in Commun. **18**, 1129 (2000).

[19] S. Geman and D. Geman, IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-6**, 721 (1984).

[20] J. G. Proakis, *Digital Communications* (McGraw-Hill, New York, 2000), 4th ed.

[21] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 3rd edition 1991).

[22] R. Chellappa and A. Jain, *Markov Random Fields : Theory and Application* (Academic Press, San Diego, CA, 1991).

[23] S. Z. Li, *Markov Random Field Modeling in Image Analysis* (Springer Verlag, Berlin, 2001), 2nd ed.

[24] J. Besag, Journal Royal Statist. Soc. B. **36**, 192 (1974).

[25] C. A. Bouman and K. Sauer, IEEE Trans. Image Processing **2**, 296 (1993).

[26] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, IEEE Trans. Image Processing **1**, 205 (1992).

[27] J. Besag, Journal of Applied Statistics **16**, 395 (1989).

[28] P. J. Green, IEEE Trans. on Medical Imaging **9**, 84 (1990).

[29] S. Geman and D. McClure, In Proc. Statist. Comput. sect. Amer. Stat. Assoc. pp. 12–18 (1985).

[30] D. Geman and G. Reynolds, IEEE Trans. on Patt. Anal. and Mach. Intell. **14**, 367 (1992).

[31] T. Hebert and R. Leahy, IEEE Trans. on Medical Imaging **8**, 194 (1989).

TABLE I: Commonly used potential functions for pair-wise cliques in generalized form where Δ is the difference between x and a neighboring pixel and p, δ are free parameters.

Name	Potential function $V_C(\Delta)$
Gauss	Δ^2
Besag [27]	$\lvert\Delta\rvert$
Green [28]	$\log\cosh(\Delta/p)$
Geman and McClure [29]	$\frac{\Delta^2}{\delta^2+\Delta^2}$
Bouman and Sauer [25]	$\lvert\Delta\rvert^p$
Geman and Reynolds [30]	$\frac{\lvert\Delta\rvert}{\delta+\lvert\Delta\rvert}$
Hebert and Leahy [31]	$\log(\delta^2 + \Delta^2)$

List of Figure Captions

Fig. 1:

Eight-pixel neighborhood system and all associated cliques.

Fig. 2:

A priori densities generated with different potential functions for neighborhood 35, 45, 65, 100. Top: smooth functions; bottom non-smooth functions. All densities are normalized as in (12). The parameters are as follows. Gauss: $T = 950$; Besag $T = 12$; Green: $T = 7$, $p = 3$; Geman and McClure: $T = 0.4$, $\delta = 10$; Bouman and Sauer: $T = 3$, $p = 0.7$; Geman and Reynolds: $T = 0.24$, $\delta = 10$; Hebert and Leahy $T = 2.8$, $\delta^2 = 0.7$.

Fig. 3:

PSNR versus the number of iterations for direct PCM transmission at $E_b/N_0 = 0$ dB.

Fig. 4:

PSNR results for transmission of the original image and different decoding methods as a function of the E_b/N_0 ratio.

Fig. 5:

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

Fig. 6:

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

Fig. 7:

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. Left: hard decoding; right: MS soft decoding.

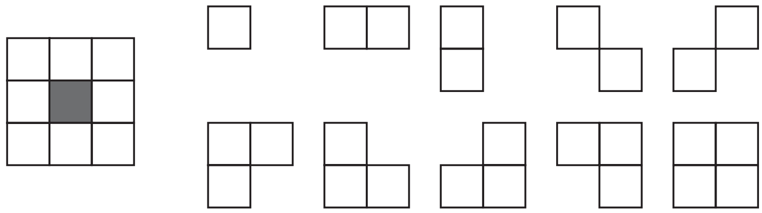

FIG. 1: Eight-pixel neighborhood system and all associated cliques.

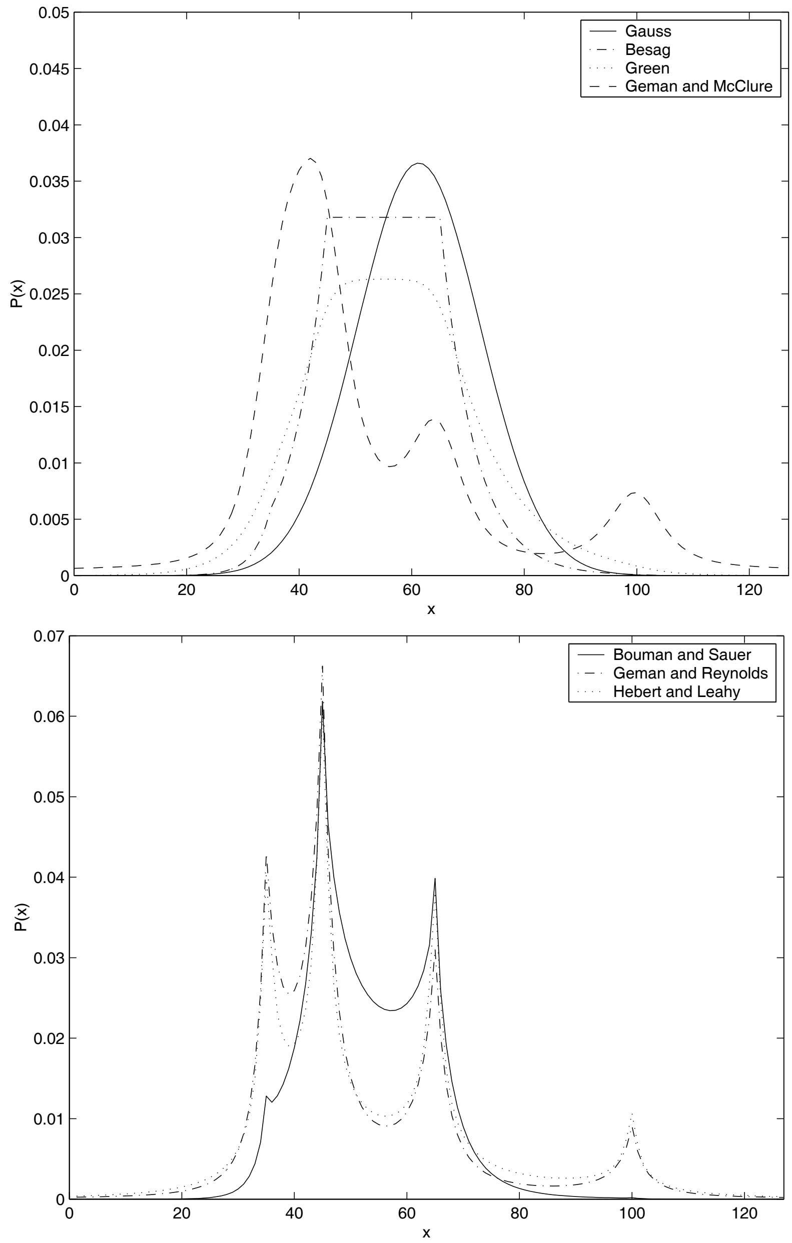

FIG. 2: A priori densities generated with different potential functions for neighborhood 35, 45, 65, 100. Top: smooth functions; bottom non-smooth functions. All densities are normalized as in (12). The parameters are as follows. Gauss: $T = 950$; Besag $T = 12$; Green: $T = 7$, $p = 3$; Geman and McClure: $T = 0.4$, $\delta = 10$; Bouman and Sauer: $T = 3$, $p = 0.7$; Geman and Reynolds: $T = 0.24$, $\delta = 10$; Hebert and Leahy $T = 2.8$, $\delta^2 = 0.7$.

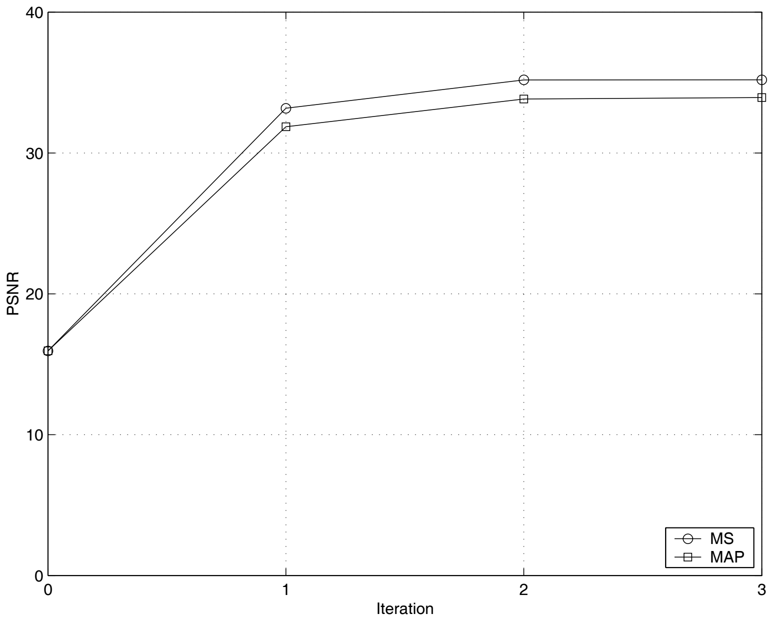

FIG. 3: PSNR versus the number of iterations for direct PCM transmission at $E_b/N_0 = 0$ dB.

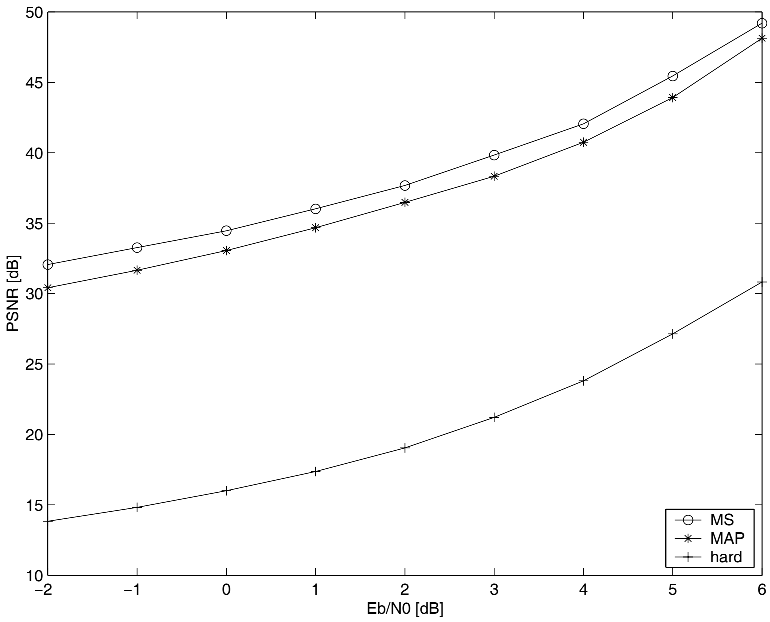

FIG. 4: PSNR results for transmission of the original image and different decoding methods as a function of the E_b/N_0 ratio.

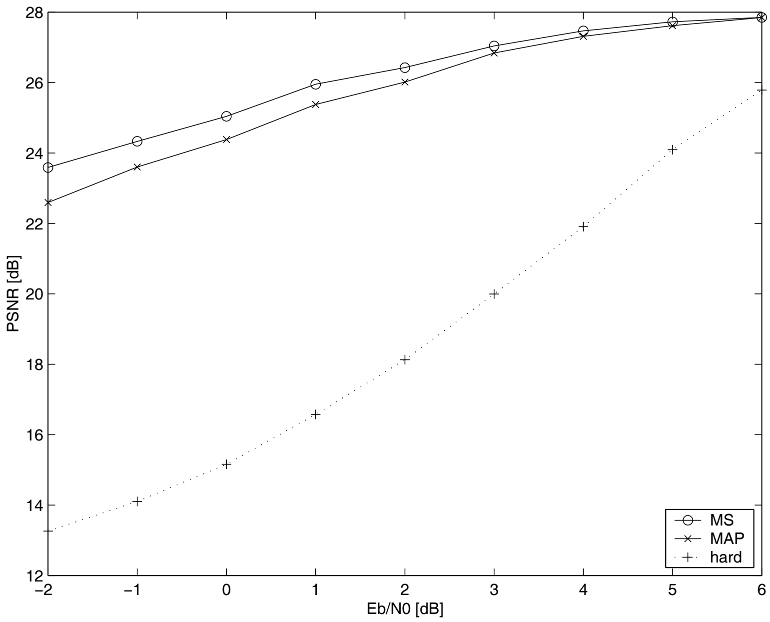

FIG. 5: PSNR results for transmission of wavelet coefficients at 0.36 bpp and different decoding methods.

FIG. 6: Examples of decoded images where the original 8 bpp image was transmitted over a channel with an E_b/N_0 of 0 dB. Left: hard decoding; right: MS soft decoding.

FIG. 7: 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. Left: hard decoding; right: MS soft decoding.