SOFT-INPUT SOURCE DECODING FOR ROBUST TRANSMISSION OF COMPRESSED IMAGES USING TWO-DIMENSIONAL OPTIMAL ESTIMATION

Jörg Kliewer and *Norbert Görtz*

University of Kiel
Institute for Circuits and Systems Theory
24143 Kiel, Germany
{jkl, ng}@tf.uni-kiel.de

ABSTRACT

In this paper we address the transmission of compressed images over highly corrupted AWGN-channels using an optimal estimation approach at the decoder. In contrast to other methods we only use a negligible amount of explicit redundancy based on channel codes. Mainly, the implicit residual source redundancy inherent in the quantized subband images and the bit-reliability information at the channel output are utilized for error protection. As a novelty we extend the optimal estimation technique from the one- to the two-dimensional case, where both horizontal and vertical correlations are exploited in the subband images. Based on this approach the performances for several estimation methods are compared, where also approaches for approximating the source correlations at the decoder are discussed.

1. INTRODUCTION

Recently, the application of joint source-channel coding for the transmission of compressed image data has attained much interest, since the classical source-channel separation scheme seems not to be justified for finite-length signals. A subset of these approaches is given by joint source-channel *de*coding, where the residual redundancy after source encoding is used in order to improve the error resilience [1–3].

In the following a joint source-channel decoding approach for image transmission over AWGN-channels is presented, where the output signal in the decoder is optimally estimated depending on bit-reliability information at the channel output and on the source statistics. In contrast to many other approaches in the literature we do not use any channel codes for error protection, except for the bit allocation. Instead only a simple wavelet-based source coder is applied, which deliberately leaves some redundancy in the source-encoded bitstream. This a-priori knowledge is then is exploited for error-concealment at the receiver. In extension to the work in [4] we here present a two-dimensional correlation model, which will subsequently be compared to one-dimensional approaches for different representations of the source redundancy at the decoder.

2. SOFT-INPUT SOURCE DECODING

2.1. Transmission system

The block diagram of the underlying transmission system is depicted in Fig. 1. Herein, the two-dimensional input image, which may for example be obtained by subband decomposition, is

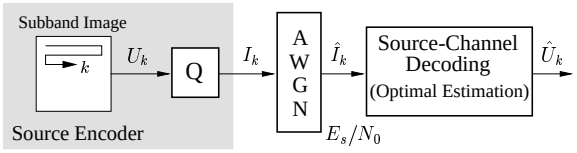

Figure 1: Model of the transmission system

scanned in order to obtain the one-dimensional source signal vector $\boldsymbol{U} = [U_0, U_1, \ldots, U_k, U_{k+1}, \ldots]$ with k denoting the sample index. The subsequent (vector-) quantization leads to indices I_k which can be represented by N bits. Due to delay and complexity constraints for the subband decomposition and the quantization we can generally assume that there is always a certain amount of residual redundancy in the vector $[I_0, I_1, \ldots, I_k, I_{k+1}, \ldots]$. The correlation of the indices is modeled by a first-order stationary Markov-process, where the Markov-model is described by the index-transition probabilities $P(I_k = \lambda \mid I_{k-1} = \mu)$, $\lambda, \mu = 0, 1, \ldots, 2^N - 1$.

In the above model the source-encoded index I_k is then transmitted over an AWGN-channel with coherently detected BPSK (Fig. 2). After a mapping to bipolar bits we add a white Gaussian noise

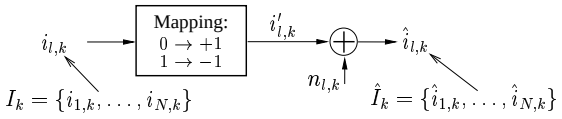

Figure 2: Transmission model for a single bit $i_{l,k}$

sample $n_{l,k}$. The noise samples have zero mean and a variance of $\sigma_n^2 = \frac{N_0}{2E_s}$, where E_s denotes the energy used to transmit each bit and N_0 represents the one-sided power spectral density of the channel noise. For the conditional p.d.f. $p(\hat{i}_{l,k}|i_{l,k})$ we then obtain

$$p\big(\hat{i}_{l,k} \mid i_{l,k}\big) = \frac{e^{-\frac{1}{2\sigma_n^2}(\hat{i}_{l,k}-1)^2}}{\sqrt{2\pi}\sigma_n} \; e^{-\frac{1}{2\sigma_n^2}\cdot 4\cdot \hat{i}_{l,k}\cdot i_{l,k}} \tag{1}$$

with $i_{l,k} \in \{0, 1\}$ and $\hat{i}_{l,k} \in \mathbb{R}$. Since we have a memoryless channel the conditional p.d.f. for the overall received soft-bit vector $\hat{I}_k$, when a certain $I_k \in \{0, 1\}^N$ is given, can be written as

$$p\big(\hat{I}_k \mid I_k\big) = \prod_{l=1}^{N} p\big(\hat{i}_{l,k} \mid i_{l,k}\big). \tag{2}$$

The source-channel decoding stage in Fig. 1 then exploits the channel-based knowledge in the $\hat{I}_k$ and the a-priori information of the source in order to maximize the signal-to-noise ratio (SNR) at the decoder output.

2.2. A-posteriori probabilities

We are now interested in the a-posteriori probabilities

$$P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k) := P(I_k = \lambda \mid \hat{I}_0, \ldots, \hat{I}_{k-1}, \hat{I}_k),$$

which denote the probability that the index $I_k = \lambda$ has been transmitted for $\lambda = 0, \ldots, 2^N - 1$, given all received soft-bit vectors $\hat{\boldsymbol{I}}_0^k = [\hat{I}_0, \ldots, \hat{I}_{k-1}, \hat{I}_k]$ up to the time instant k. The a-posteriori probabilities represent a reliability information for the hypothesis $I_k = \lambda$, for which also the notation $I_k^{(\lambda)}$ will be used in the sequel.

With the index transition probabilities $P(I_k = \lambda \mid I_{k-1} = \mu)$ of the Markov model and the conditional p.d.f.s in (1) and (2), resp., the a-posteriori probabilities $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ can be calculated with the recursion [3, 5]

$$P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k) = c_k \cdot p\big(\hat{I}_k \mid I_k = \lambda\big) \cdot \cdot \sum_{\mu=0}^{2^N-1} P\big(I_k = \lambda \mid I_{k-1} = \mu\big)\; P(I_{k-1}^{(\mu)} \mid \hat{\boldsymbol{I}}_0^{k-1}). \quad (3)$$

The factor $c_k \in \mathbb{R}$ ensures that the $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ are true probabilities, and the initialization for $k = 0$ can be carried out with the unconditional source-index probabilities $P(I_k = \lambda)$.

2.3. Optimal estimation

The a-posteriori probabilities obtained in (3) can now be used for optimally estimating the source-encoded indices U_k in such a way that the value of a "suitable" overall distortion is minimized in the reconstructed vector $\hat{\boldsymbol{U}} = [\hat{U}_0, \hat{U}_1, \ldots, \hat{U}_k, \hat{U}_{k+1}, \ldots]$ at the decoder output.

(a) Mean-squares estimation. The mean-squares (MS) estimator is especially well suited for waveform-like signals, since it directly corresponds to the demand for maximal SNR. A MS-estimation can be obtained according to

$$\hat{U}_k^{(\mathrm{MS})} = \sum_{\lambda=0}^{2^N-1} U_q(\lambda) \cdot P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k), \quad (4)$$

where each entry $U_q(\lambda)$ of the quantization table is weighted with the a-posteriori probability corresponding to the index λ prior to summation. Furthermore, this estimator has a "graceful degradation" property. This can be best seen from the worst case example, when all $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ are equally distributed in (4). In this case the estimated value $\hat{U}_k^{(\mathrm{MS})}$ takes on the mean of $U_q(\lambda)$ for all λ.

(b) Maximum a-posteriori estimation. The maximum a-posteriori (MAP) estimator minimizes the decoding error probability in the reconstructed values, where this estimator is defined as

$$\hat{U}_k^{(\mathrm{MAP})} = U_q(I_k = \lambda_{\mathrm{map}}), \;\; P(I_k^{(\lambda_{\mathrm{map}})} \mid \hat{\boldsymbol{I}}_0^k) \geq P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k). \quad (5)$$

A special case occurs when the source indices are assumed to be independent and equally distributed, that is $P(I_k = \lambda | I_{k-1} = \mu) = P(I_k = \lambda) = 1/2^N$ for all μ, λ. Then no a-priori information is available, which corresponds to classical maximum likelihood or hard decision decoding at the channel output.

2.4. Extension to two-dimensions

(a) Problem statement. Up to now we have only regarded the source-index correlations in one dimension. However, since many subband images have spatial correlations in the horizontal *and* vertical direction, this additional knowledge can be exploited by two-dimensional a-posteriori probabilities.

Due to complexity reasons in the following we assume a relevant correlation only between spatially adjacent pixels. This is depicted in Fig. 3, where some received indices of a subband image are displayed, and $\hat{\boldsymbol{I}}_0^M = [\hat{I}_0, \hat{I}_1, \ldots, \hat{I}_M]$ denotes a received row or column vector of length $M + 1$. The a-priori knowledge

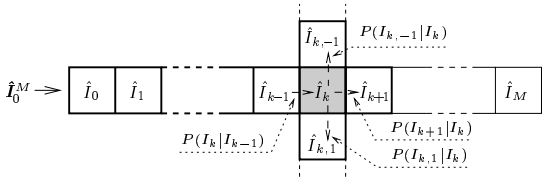

Figure 3: Received values in a subband image, and corresponding Markov models (see text)

is restricted to all indices in the boxes surrounded with bold lines, namely the nearest neighbors of $\hat{I}_k$ in the two-dimensional subband image and all already received values in horizontal or vertical direction. Thus, the a-posteriori probabilities can now be written as $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1})$, for which in the following an expression based on the probabilities $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ in (3), the channel p.d.f.s in (1) and (2) and the transition probabilities $P(I_{k+1}^{(\mu)} \mid I_k^{(\lambda)})$, $P(I_{k,1}^{(\nu)} \mid I_k^{(\lambda)})$ and $P(I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)})$ of the Markov sources in Fig. 3 will be derived.

(b) Derivation of the two-dimensional a-posteriori probability. In order to simplify the derivation, we first state the following relation for the conditional p.d.f.s with the random variables $a, b, c \in \mathbb{R}$ [6][1]:

$$p(a, b \mid c) = \frac{p(b, c)}{p(c)} \cdot \frac{p(a, b, c)}{p(b, c)} = p(b \mid c) \cdot p(a \mid b, c) \quad (6)$$

In a first step, the a-posteriori probabilities can be written with (6) according to

$$P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1}) = c'_k \cdot p(I_k^{(\lambda)}, \hat{I}_{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1} \mid \hat{\boldsymbol{I}}_0^k) = c'_k \cdot p(\hat{I}_{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1} \mid I_k^{(\lambda)}) \cdot P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k), \quad (7)$$

where in the last equation both the memoryless property of the channel and the Markov property of the source are utilized. The factor c'_k is independent of λ and ensures that the expression on the left-hand side of (7) is a true probability. The joint p.d.f. $p(\hat{I}_{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1} \mid I_k^{(\lambda)})$ can now be further decomposed for a memoryless channel and by use of (6) as

$$p(\hat{I}_{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1} \mid I_k^{(\lambda)}) = \sum_{\mu=0}^{2^N-1} \sum_{\nu=0}^{2^N-1} \sum_{\kappa=0}^{2^N-1} p(\hat{I}_{k+1} \mid I_{k+1}^{(\mu)}) \cdot \cdot p(\hat{I}_{k,1}, \hat{I}_{k,-1}, I_{k+1}^{(\mu)}, I_{k,1}^{(\nu)}, I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)}). \quad (8)$$

[1] Note that the same relation holds for discrete random variables, where the p.d.f.s are then replaced by probabilities.

Similarly, the two other channel terms $p(\hat{I}_{k,1} \mid I_{k,1}^{(\nu)})$ and $p(\hat{I}_{k,-1} \mid I_{k,-1}^{(\kappa)})$ can be extracted from the joint conditional p.d.f. on the right-hand side in equation (8), leading to

$$p(\hat{I}_{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1} \mid I_k^{(\lambda)}) = \sum_{\mu=0}^{2^N-1} \sum_{\nu=0}^{2^N-1} \sum_{\kappa=0}^{2^N-1} p(\hat{I}_{k+1} \mid I_{k+1}^{(\mu)}) \cdot \cdot p(\hat{I}_{k,1} \mid I_{k,1}^{(\nu)})\, p(\hat{I}_{k,-1} \mid I_{k,-1}^{(\kappa)})\, P(I_{k+1}^{(\mu)}, I_{k,1}^{(\nu)}, I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)}). \quad (9)$$

The term $P(I_{k+1}^{(\mu)}, I_{k,1}^{(\nu)}, I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)})$ describes the joint transition probabilities from the actual index I_k under consideration to all possible combinations of the indices I_{k+1} and $I_{k,\pm1}$. However, in order to store this joint probability a four-dimensional matrix is needed, leading to high memory requirements especially for larger N. For example, for $N = 8$ we would need $2^{4N} \approx 4 \cdot 10^9$ storage cells, which is practically infeasible. However, a simplification is possible by expressing the joint probability with a product of separate terms:

$$P(I_{k+1}^{(\mu)}, I_{k,1}^{(\nu)}, I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)}) \approx P(I_{k+1}^{(\mu)} \mid I_k^{(\lambda)})\, P(I_{k,1}^{(\nu)} \mid I_k^{(\lambda)})\, P(I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)}). \quad (10)$$

Note that this is only an approximation, since for natural subband images there are also some diagonal correlations present between the indices $I_{k,\pm1}$ and I_{k+1} (Fig. 3).

By combining the equations (7), (9), and (10) we now obtain the final expression for the a-posteriori probabilities as

$$P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1}) \approx c_k' \cdot P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k) \cdot \cdot \sum_{\mu=0}^{2^N-1} p(\hat{I}_{k+1} \mid I_{k+1}^{(\mu)})\, P(I_{k+1}^{(\mu)} \mid I_k^{(\lambda)}) \cdot \cdot \sum_{\nu=0}^{2^N-1} p(\hat{I}_{k,1} \mid I_{k,1}^{(\nu)})\, P(I_{k,1}^{(\nu)} \mid I_k^{(\lambda)}) \cdot \cdot \sum_{\kappa=0}^{2^N-1} p(\hat{I}_{k,-1} \mid I_{k,-1}^{(\kappa)})\, P(I_{k,-1}^{(\kappa)} \mid I_k^{(\lambda)}). \quad (11)$$

We can see that in order to obtain the two-dimensional a-posteriori probabilities the reliability values $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ from (3) are weighted with three sum-terms. Each sum contains the corresponding transition probability and the channel term for the "new" indices I_{k+1}, $I_{k,1}$ and $I_{k,-1}$, respectively.

MS and MAP estimation at the decoder can again be carried out by simply replacing the probabilities in (4) and (5), resp., with the values $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1})$ from (11).

3. ROBUST TRANSMISSION OF COMPRESSED IMAGES

The optimal estimation approach from Section 2.3 is now applied to an experimental image codec, where the overall transmission system is shown in Fig. 5. For the subband decomposition we utilize a wavelet-based octave filter bank with L levels and the well-known 9-7 subband filters [7]. The bit-allocation is carried out in a rate-distortion optimal sense [8], where the subband signals are quantized with simple scalar quantizers. Since the bit-allocation information is highly sensitive to channel errors, we assume that this information is protected by a sufficiently strong channel code and thus is transmitted without errors. At the decoder side, the optimal estimation of the reconstructed subband coefficients $\hat{U}_k^{(\ell)}$ is carried out independently for every subband image.

Prior to quantization the two-dimensional subband images have to be scanned in order to obtain an one-dimensional subband vector $\boldsymbol{U}^{(\ell)} = [U_0^{(\ell)}, U_1^{(\ell)}, \ldots, U_k^{(\ell)}, \ldots]$, $\ell = 0, \ldots, K-1$ with $K = 3\,(L-1)+4$. The scanning is carried out in a meander-type fashion depicted in Fig. 4 for the first level of the decomposition [4]. Thus the spatial correlations in scanning direction are projected onto correlations between the elements of $\boldsymbol{U}^{(\ell)}$. Also, we choose the scanning orientation such that the correlations between the elements of $\boldsymbol{U}^{(\ell)}$ are maximized. In the HL- and LH-subband this corresponds with the orientation of the low-pass filtering, where in the LL- and HH-subbands the orientation can be chosen arbitrarily.

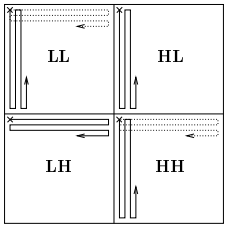

Figure 4: Scanning of the different subband images

4. RESULTS

The experimental image transmission system is applied to the "Goldhill" test image of pixel dimension 512×512 for a $L = 3$ level decomposition and a source coding rate of 0.36 bit per pixel (bpp) including all side information. We compare different estimation techniques, where the MS/MAP estimation is carried out with the following a-posteriori probabilities:

- $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ from (3) for all subbands, denoted as "1D, P".
- $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1})$ for all subbands, which can be calculated from (11) by omitting the sum-terms for $I_{k,1}$ and $I_{k,-1}$ ("1D, PF1").
- $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^{k+1}, \hat{I}_{k,-1}, \hat{I}_{k,1})$ from (11) for the LL-subband and the lowest LH-, HL- and HH-subbands, $P(I_k^{(\lambda)} \mid \hat{\boldsymbol{I}}_0^k)$ otherwise ("2D"). This approach has been chosen, since only for the lower subbands there is both horizontal and vertical correlation inherent in the subband images, such that an application of the 2D a-posteriori probability also to the upper subbands does not lead to a major improvement.

Furthermore, different techniques for obtaining the index transition probabilities at the decoder are utilized. In the optimal case these probabilities are directly derived from the original image, which will be denoted with "Orig." in the following. However, this method is infeasible in real transmission scenarios, which require approximations for the true transition probabilities to be calculated at the decoder. In the following we address two approximation methods:

- The transition probabilities are computed from a training set of 129 images (i.e. faces, landscapes, satellite images), where the "Goldhill" test image is not included ("Tr.").
- The transition probabilities are derived by two-dimensional AR(1) modelling from the original image ("AR(1)").

In Fig. 6 the average peak SNR (PSNR) values of the reconstructed "Goldhill" image versus the channel-SNR E_s/N_0 are displayed for different estimation techniques, where for each value of E_s/N_0 50 independent trials have been simulated. When we only regard