pass through the AWG twice before reaching detectors, thus similarly suppressing crosstalk between the chips. This very selective chip filtering characteristic improves the ability of the decoder to reject noise and multiple access interference [4, 8].

Summary: We have presented an integrated planar lightwave circuit design to implement a compact CSK encoder/decoder. This design has several advantages. First, the use of a single AWG eliminates the problem of spectral misalignment due to fabrication errors in a multiple AWG configuration. Secondly, the reflective structure reduces the size and number of components, along with the power requirements for switches and phase shifters. Thirdly, the switching contrast and power of each spectral chip can be maximised and equalised using the power and phase controls in each arm of the MZI to compensate for source, coupler, and other component characteristics. Finally, the programmable encoding and decoding masks allow users to have flexible and secure accessibility to the networks. An AWG with a 25GHz channel spacing in this design would permit code patterns having 200 chips within the spectrum of an Er-doped superfluorescent fibre source. The compactness of the proposed device makes mass production and commercial availability feasible at reasonable costs.

6 August 1999
Electronics Letters Online No: 19991263
DOI: 10.1049/el:19991263

Chau-Han Lee, Shan Zhong, Xiao Lin, and Y.J. Chen (*Department of Computer Sciences and Electrical Engineering, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA*)

J.F. Young (*Department of Electrical and Computer Engineering, Rice University, MS-366, PO Box 1892, Houston, TX 77251, USA*)

E-mail: clee@gl.umbc.edu

References

1 SALEHI, J.A., WEINER, A.M., and HERITAGE, J.P.: 'Coherent ultrashort light pulse code-division multiple access communication systems', *J. Lightwave Technol.*, 1990, **8**, (3), pp. 478–491

2 GRIFFIN, R.A., SAMPSON, D.D., and JACKSON, D.A.: 'Coherence coding for photonic code-division multiple access communication networks', *J. Lightwave Technol.*, 1995, **13**, (9), pp. 1862–1837

3 SARDESAI, H.P., CHANG, C.C., and WEINER, A.M.: 'A femtosecond code-division multiple access communication system test bed', *J. Lightwave Technol.*, 1998, **16**, (11), pp. 1953–1964

4 KAVEHRAD, M., and ZACCARIN, D.: 'Optical code-division-multiplexed systems based on spectral encoding of noncoherent sources', *J. Lightwave Technol.*, 1995, **13**, (3), pp. 534–545

5 ADAM, L., SIMOVA, E., and KAVEHRAD, M.: 'Experiment on optical code-division-multiple-access switch system using spectral amplitude encoding of light-emitting diodes (Applications of Photonic Technology 2)' (Plenum Press, 1997), pp. 379–384

6 ZACCARIN, D., and KAVEHRAD, M.: 'Performance evaluation of optical CDMA systems using non-coherent detection and bipolar codes', *J. Lightwave Technol.*, 1994, **12**, (1), pp. 96–105

7 KARAFOLAS, N., and UTTAMCHANDANI, D.: 'Optical CDMA system using bipolar codes and based on narrow passband optical filtering and direct detection', *IEEE Photonics Technol. Lett.*, 1995, 7, (9), pp. 1072–1074

8 NGUYEN, L., DENNIS, T., AAZHANG, B., and YOUNG, J.F.: 'Experimental demonstration of bipolar codes for optical spectral amplitude CDMA communication', *J. Lightwave Technol.*, 1997, **15**, (9), pp. 1647–1653

9 DENNIS, T., and YOUNG, J.F.: 'Measurements of BER performance for bipolar encoding of an SFS', to be published in *J. Lightwave Technol.*, September 1999

10 LAM, C.F., TONG, D.T.K., WU, M.C., and YABLONOVITCH, E.: 'Experimental demonstration of bipolar optical CDMA system using a balanced transmitter and complementary spectral encoding', *IEEE Photonics Technol. Lett.*, 1998, **10**, (10), pp. 1504–1506

11 OGUCHI, K.: 'New notations based on the wavelength transfer matrix for functional analysis of wavelength circuits and new WDM networks using AWG-based star coupler with asymmetric characteristics', *J. Lightwave Technol.*, 1996, **14**, (6), pp. 1255–1263

Rate control strategy for embedded wavelet video coders

R. Caetano and E.A.B. da Silva

An investigation into the use of rate-control strategies for embedded wavelet video encoders is presented. It is shown that the best model for the R-D of these encoders is piecewise linear. Also, an effective iterative procedure is proposed for dealing with the problem of frame dependency, which yields improved rate × distortion results.

Introduction: Embedded wavelet coders have been very successful in the coding of still images. Nevertheless, for interframe video these encoders often have not been as successful as expected, and classical DCT-based methods are in general preferred. However, this class of video encoder is still being actively researched; for example, good wavelet video encoding results have been recently reported in [1]. Among the chief reasons for this interest in wavelet video encoders is their excellent performance for still images together with their ability to achieve precise control over the bit rate of every frame, which is well suited for constant bit rate (CBR) applications.

The simplest solution for the bit-rate allocation in these encoders is to divide the bit-budget of a sequence equally among all its frames. This solution would in principle provide the advantage of obviating the need for a buffer in order to smooth out bit-rate variations. In this Letter we investigate alternative forms of bit allocation aiming at obtaining near-optimal solutions in terms of average signal-to-noise ratio over the entire sequence. Unlike previous work [2], in which it was assumed that the rate × distortion (R-D) curve for a single frame difference is exponential, we begin by showing that the best model for the R-D characteristics of an embedded wavelet video encoder is piecewise linear. This enables us to perform rate allocation using Lagrangian optimisation at low computational complexity. We then propose an iterative procedure for dealing with the problem of frame dependency. Simulation results show that the proposed rate-control strategy consistently improves the R-D performance of embedded wavelet-based video encoders.

Implementation details: In this work we employ two types of embedded wavelet encoders. The EZW [3], which is based on successive approximation scalar quantisation, and the SA-W-VQ, which is based on successive approximation vector quantisation [4]. In both coders, we use a two-stage biorthogonal wavelet transform as recommended in [4]. In the SA-W-VQ case, blocks of wavelet coefficients of dimension 2 × 2, 2 × 4 and 4 × 4 are encoded employing the first shells of the root lattices D_4, E_8 and Λ_{16}, respectively.

The video encoder used in this implementation is based on the MPEG-4 VM-8 [5]. The main modification to it is that the DCT, quantisation and run-length encoding followed by Huffman coding have been replaced by an embedded wavelet encoder followed by an adaptive arithmetic coder. Another important modification has to do with the fact that the wavelet transform is no longer applied to independent image blocks, but to the image as a whole. The image is divided in macroblocks only for the purposes of motion estimation and compensation. Therefore, in a coded frame, either all of its macroblocks are intra-frame or all of its macroblocks are inter-frame (restricted to forward prediction, P frames). In addition, the advanced prediction mode [5] is always turned on. We also consider each frame as composed of only one rectangular VOP, identical to the frame itself. It is important to point out that, to take into account in the performance evaluation only the influence of the interframe-encoding algorithms, the first frame of a sequence was never encoded in the simulations.

Rate control: The problem of allocating the rate among the frames of a video encoder is that of, given a set of N frames $\{f_i, i = 1, ..., N\}$ and an average target rate of R_{target} bits/frame, encoding frame f_i with rate R_i, yielding distortion D_i, such that the average rate R is less than or equal to R_{target} and the average distortion D is minimum.

Considering that the rate-distortion (R-D) functions of each frame f_i are convex, this problem can be solved via Lagrangian optimisation, being equivalent to minimising the functional $D + \lambda R$. If we make the extra assumption that the R-D characteristics of frame f_i are independent of the particular point (R_j, D_j) in which frame j is being coded, $\forall j \neq i$, then this problem can be further simplified. It becomes equivalent to minimising $D_i + \lambda R_i$, for each i, subject to the constraint on the average rate [6].

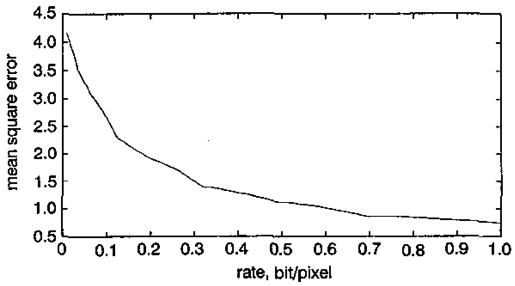

Fig. 1 *R-D characteristics of interframe prediction error encoded by embedded wavelet encoder*

In [2], Cheng *et al.* investigated the use of an EZW-like encoder in interframe coding, and developed an optimum rate allocation strategy based on the supposition that the R-D characteristics of each frame are exponential. However, we claim here that, in embedded wavelet encoders, the R-D characteristics of each frame are better represented by a piecewise linear model. This claim is based on the fact that the wavelet coefficients in such encoders are transmitted by bit-planes [Note 1]. Within a certain bit-plane, for each bit sent, there is a fixed reduction in the total distortion, corresponding to the precision associated with that bit-plane. Therefore, for a given bit-plane, the R-D characteristics are linear. In the next bit-plane, the precision is increased, and the absolute slope of the R-D characteristics decreases. Such behaviour gives rise to piecewise-linear characteristics. This can be verified by Fig. 1, which shows the R-D characteristics of the motion compensated frame difference between frames 003 and 000 of the mother-and-daughter sequence when encoded by an SA-W-VQ encoder [4] using E_8 as the orientation codebook and an α of 0.55. It is important to note that an exponential model will only fit an R-D curve if the number of bit-planes is large, which is seldom the case for interframe coding, specially in low bit rate applications. Also, an important advantage of the piecewise-linear model is that it allows a large reduction in the complexity of the Lagrangian optimisation [7].

In interframe video encoders, the prediction error corresponding to frame f_i in general depends on how frame f_i.1 has been encoded. Therefore, for each (R_i, D_i) point in frame f_i, there will be a different R-D characteristic for frame f_{i+1}. This violates the frame independence assumption made above, leading to suboptimal results. It is called the 'dependency problem' [8]. In this Letter, we propose an iterative procedure for dealing with it. Once the reconstructed frames for iteration $n - 1$ are obtained, the rate allocation for iteration n is computed, and then the reconstructed frames for iteration n can be obtained. This process proceeds until either the change in the signal-to-noise ratio performance is below a threshold or a maximum number of iterations is exceeded. In our simulations, this process has led to an increase in signal-to-noise ratio performance in all cases considered. It should be noted that one advantage of this approach is that, instead of devising arbitrary models for the frame dependency, as in [2], we have only resorted to the minor assumption that each iteration of the process leads to an improvement in the signal-to-noise ratio.

Experimental results: We have coded the sequences mother-and-daughter, silent and hall-monitor with 300, 450 and 330 QCIF frames at 30 frame/s, subsampled in time by a factor of 3 to generate 10 frame/s sequences. The bit rate in all experiments was 64kbit/s.

Table 1 shows a comparison of the average peak signal to noise ratio (PSNR) of the different MPEG-4 adaptations. It can be seen from this Table that the proposed rate-control scheme consistently improves the rate-distortion results when compared to the constant-rate case, for both EZW and SA-W-VQ (using lattices D_4, E_8 and Λ_{16}). Also, using the proposed scheme, the EZW performance has been increased so that it became competitive with that of the DCT.

Table 1: Comparison between average PSNR [dB] of different MPEG-4 variations

	Constant rate			Proposed rate control		
	Mother	Silent	Hall	Mother	Silent	Hall
DCT	–	–	–	39.26	35.77	39.30
EZW	38.25	36.04	38.49	38.70	36.49	39.31
Λ_{16}	39.61	37.52	39.80	39.90	37.56	40.07
E_8	39.46	37.32	39.82	39.71	37.46	40.15
D_4	39.22	37.09	39.79	39.48	37.32	40.11

'Rate-control' means proposed strategy using three iterations in group of 40 frames; in DCT case, MPEG-4 VM-8 is used

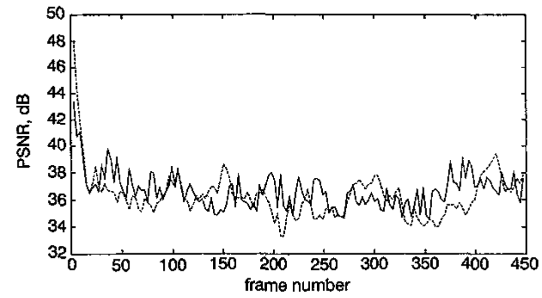

Fig. 2 *PSNR against frame number for silent sequence, coded using EZW*

——— rate control
– – – – constant rate

Fig. 2 shows the PSNR plotted against frame number for the wavelet-coded sequence 'silent', with and without rate control.

Conclusions: We have proposed a novel rate-control strategy for use in embedded wavelet video encoders which leads to improved PSNR results. We have shown that the piecewise linear model is the best model for the R-D characteristics of the frame difference encoded using such coders. The frame dependency problem has been tackled by iteratively applying the proposed strategy to a group of frames.

25 August 1999
Electronics Letters Online No: 19991272
DOI: 10.1049/el:19991272

R. Caetano and E.A.B. da Silva (*Universidade Federal do Rio de Janeiro, Cx. P. 68504, Rio de Janeiro, RJ, 21945-970, Brazil*)

E-mail: caetano@lps.ufrj.br

References

1 MARPE, D., and CYCON, H.L.: 'Very low bit-rate video coding using wavelet-based techniques', *IEEE Trans. Circuits Syst. Video Technol.*, 1999, **9**, pp. 85–94

2 LI, J., CHENG, P.-Y., and KUO, C.-C.J.: 'Rate control for an embedded wavelet video coder', *IEEE Trans. Circuits Syst. Video Technol.*, 1997, **7**, pp. 696–702

3 SHAPIRO, J.M.: 'Embedded image coding using zerotrees of wavelet coefficients', *IEEE Trans. Acoust. Speech Signal Process.*, 1993, **ASSP-41**, pp. 3445–3462

4 SAMPSON, D.G., DA SILVA, E.A.B., and GHANBARI, M.: 'Low bit rate video coding using wavelet vector quantisation', *IEE Proc. I*, 1995, **142**, pp. 141–148

5 ISO/IEC JTC1/SC29/WG11, 'MPEG-4 video verification model version 8.0'. July 1997

Note 1: In the SA-W-VQ algorithm the set of code vectors in a given pass is considered as a 'vector bit-plane' [4]

6 RAMCHANDRAN, K., ORTEGA, A., and VETTERLI, M.: 'Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders', *IEEE Trans. Image Process.*, 1994, **IP-3**, pp. 533–545

7 LIN, L.-J., and ORTEGA, A.: 'Bit-rate control using piecewise approximated rate-distortion characteristics', *IEEE Trans. Circuits Syst. Video Technol.*, 1998, **8**, pp. 446–459

8 RAMCHANDRAN, K., and ORTEGA, A.: 'Rate-distortion methods for image and video compression', *IEEE Sig. Process.*, 1998, **SP-15**, pp. 23–50

Speaker adaptation technique for HMM model

S. Kwong, Q.H. He, K.F. Man and K.S. Tang

A new speaker adaptation technique for the hidden Markov model (HMM) based on the maximum model distance (Kwong, 1998; He, 1999) approach is presented. Experimental results have shown that this technique provides good performance even with a small amount of adaptation data. When these results are compared with those from the Baum-Welch approach and the stochastic matching approaches (Sank, 1996), it is found that the presented approach provides the best performance.

Adaptation with Baum-Welch algorithm: Most hidden Markov model (HMM) adaptation techniques [1, 2] are based on approaches different from those for the estimation of HMM model parameters with abundant speech data. It is believed that those well-developed training approaches could not achieve robust results if there is only a small amount of speech data. In this Letter, we propose a new technique that can be applied to HMM adaptation in the case of sparse adaptation speech data.

An N-state HMM with continuous mixture density distribution is usually expressed by parameter vector $\lambda = (\pi, \mathbf{A}, \theta)$, $\theta_j = \{c_{jk}, \mathbf{u}_{jk}, \mathbf{R}_{jk}\}$, $k = 1, 2, ..., K$ for state j. This representation does not show the amount of training data used in the training process. Let $\mathbf{O} = \{O_c, c = 1, 2, ... C\}$ represent the training data; under the framework of the Baum-Welch algorithm, we could express λ in form $\tilde{\lambda} = (\tilde{\pi}, \tilde{\mathbf{A}}, \tilde{\theta})$, where

$$\tilde{\pi}_i = \sum_{c=1}^{C} \sum_{j=1}^{N} \zeta_1^{(c)}(i,j) \qquad (1a)$$

$$\tilde{a}_{ij} = \sum_{c=1}^{C} \sum_{t=1}^{T_c-1} \zeta_t^{(c)}(i,j) \qquad (1b)$$

$$\tilde{c}_{jk} = \sum_{c=1}^{C} \sum_{t=1}^{T_c} r_t^{(c)}(j,k) \qquad (1c)$$

$$\tilde{\mathbf{u}}_{jk} = \sum_{c=1}^{C} \sum_{t=1}^{T_c} r_t^{(c)}(j,k)\mathbf{o}_t^{(c)} \qquad (1d)$$

$$\tilde{\mathbf{R}}_{jk} = \sum_{c=1}^{C} \sum_{t=1}^{T_c} r_t^{(c)}(j,k)\left(\mathbf{o}_t^{(c)} - \mathbf{u}_{jk}\right)\left(\mathbf{o}_t^{(c)} - \mathbf{u}_{jk}\right)' \qquad (1e)$$

$\xi_t, (i, j) = P(q_t = i, q_{t+1} = j|\mathbf{O}, \lambda)$, and $\gamma_t (j, k)$ is the probability of being in state j at time t with the kth mixture component accounting for $\mathbf{o}_t$, which could be computed easily with a forward-backward procedure.

The transformation of $\tilde{\lambda}$ to λ is very simple. Let $\tilde{\lambda}_0$ be the original model parameters of a word, $\mathbf{O}^A = \{O_c^A, c = 1, 2, ... C_A\}$ be the adaptation speech data, and λ be the corresponding normal version of $\tilde{\lambda}$. The model λ_0 could be adapted as follows:

1. Estimate the contribution $\tilde{\lambda}^A$ of $\mathbf{O}^A$ through eqn. 1 based on $\lambda^{(p)}$, where p is the recursion index.
2. Define $\tilde{\lambda}^{(p+1)} = \tilde{\lambda}_0 + \tilde{\lambda}^A$.
3. Investigate the convergence of $\lambda^{(p+1)}$. If the adaptation procedure converges, take $\tilde{\lambda}^{(p+1)}$ as the final result and exit. Otherwise, let $p = p + 1$ and go to step 1 to begin the next adaptation. In fact, the above adaptation procedure shares the same framework of the Baum-Welch algorithm except for step 2. Furthermore, this approach has the following features:

(i) All parameters of λ are adapted by using the adaptation speech data $\mathbf{O}^A$. Transformation-based adaptation approaches can only adapt the means, or the means and covariance of a mixture density function [2, 3].
(ii) When $C_A = 0$, i.e. there are no adaptation data for model λ, then λ does not change.
(iii) When C_A becomes very large, or the amount of adaptation data is much larger than that used in the original model λ_0 i.e. $C_A >> C$, the final adapted system approaches an environment-dependent system. This is one of the most important features of MAP-based adaptation methods [1].
(iv) No assumption is made. However, some assumptions should be made for transformation-based and MAP-based approaches. For instance, MAP-based approaches [1] assume the joint prior density of the mixture gain, mean vector and covariance matrix of the HMM parameters to be the product of Dirichlet and normal-Wishart densities. To emphasise the effects of sparse adaptation data, the above adaptation procedure could be iterated several times.

Adaptation with maximum model distance (MMD) algorithm: The Baum-Welch algorithm does not adapt those models without a speech sample. However, the MMD approach [4, 5] could solve this problem. Let $O^v = \{O_c^v, c = 1, 2, ..., C_v\}$ be the training data labelled to model λ_v; λ_v could be equivalently expressed as $\tilde{\lambda}_v = (\tilde{\pi}\ \tilde{\mathbf{A}}^v, \tilde{\theta}^v)$.

$$\tilde{\pi}_i^v = \sum_{c=1}^{C_v} \gamma_1^{vc}(i) - \sum_{\theta-1,\theta\neq v}^{M} \frac{\max(1, C_v)}{C_\theta} \sum_{c=1}^{C_\theta} R_{\theta c}\gamma_1^{\theta c}(i) \qquad i = 1, 2, ..., N \quad (2a)$$

$$\tilde{a}_{ij}^v = \sum_{c=1}^{C_v} s_{ij}^{vc} - \sum_{\theta=1,\theta\neq v}^{M} \frac{\max(1, C_v)}{C_\theta} \sum_{c=1}^{C_\theta} R_{\theta c}^v s_{ij}^{\theta c} \qquad i,j = 1, 2, ..., N \quad (2b)$$

$$\tilde{c}_{jk}^v = \sum_{c=1}^{C_v} \gamma^{vc}(j,k) - \sum_{\theta-1,\theta\neq v}^{M} \frac{\max(1, C_v)}{C_\theta} \sum_{c=1}^{C_\theta} R_{\theta c}^v \gamma^{\theta c}(j,k) \qquad (2c)$$

$$\tilde{\mathbf{u}}_{jk}^v = \sum_{c=1}^{C_v} \mathbf{u}_{jk}^{vc} - \sum_{\theta=1,\theta\neq v}^{M} \frac{\max(1, C_v)}{C_\theta} \sum_{c=1}^{C_\theta} R_{\theta c}^v \mathbf{u}_{jk}^{\theta c} \qquad (2d)$$

$$\tilde{\mathbf{R}}_{jk}^v = \sum_{c=1}^{C_v} \mathbf{R}_{jk}^{vc} - \sum_{\theta=1,\theta\neq v}^{M} \frac{\max(1, C_v)}{C_\theta} \sum_{c=1}^{C_\theta} R_{\theta c}^v \mathbf{R}_{jk}^{\theta c} \qquad (2e)$$

where

$$\gamma_1^{\theta c}(i) = P\left(q_1 = i|O_c^\theta, \lambda_c\right)$$

$$s_{ij}^{\theta c} = \sum_{t=1}^{T_c^\theta - 1} \xi_t^{\theta c}(i,j)$$

$$R_{\theta c}^v = \frac{P^\eta(O_c^\theta|\lambda_v)}{\sum_{k=1,k\neq\theta}^{M} P^\eta(O_c^\theta|\lambda_k)}$$

$$\gamma^{\theta c}(j,k) = \sum_{t=1}^{T_c^\theta} \gamma_t^{\theta c}(j,k)$$

$$\mathbf{u}_{jk}^{\theta c} = \sum_{t=1}^{T_c^\theta} \gamma_t^{\theta c}(j,k)\mathbf{o}_t^\theta$$

$$\mathbf{R}_{jk}^{\theta c} = \sum_{t=1}^{T_c^\theta} \gamma_t^{\theta c}(j,k)\left(\mathbf{o}_t^\theta - \mathbf{u}_{jk}^\theta\right)\left(\mathbf{o}_t^\theta - \mathbf{u}_{jk}^\theta\right)'$$

Let $\tilde{\Lambda}^{(0)} = \tilde{\lambda}_i^{(0)}, i = 1, 2, ..., M\}$ be the original model set and $O^A = \{\mathbf{O}_i^A, i = 1, 2, ..., M\}$ be the adaptation speech data. The model set $\Lambda^{(0)}$ can be adapted as follows.
(i) Estimate the contribution $\tilde{\Lambda}^A$ of O^A through eqns. 2*a–e* based on the current model set $\tilde{\Lambda}^{(p)}, p = 0, 1, 2, ...,$.
(ii) Define $\tilde{\Lambda}^{(p+1)} = \tilde{\Lambda}^{(0)} + \tilde{\Lambda}^A$.
(iii) To check the convergence condition for $\Lambda^{(p+1)}$, if the adaptation procedure converges, take $\tilde{\Lambda}^{(p+1)}$ as the final result and exit. Otherwise, let $p = p + 1$ and go to step 1 to begin another adaptation recursion. It can be seen that the above MMD-based adaptation procedure preserves most of the characteristics of the Baum-