Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49(63), Fascicola 1, 2004

Video stream compression without A/D conversion using Analog Multilayer Perceptron

Spiridon Florin Beldianu[1]

Abstract –This work proposes a multilayer perceptron (MLP) method to perform direct mapping from an analog frame block into a vector index, implement vector quantization (VQ) without the need for carrying out first an analog-to-digital (A/D) conversion of each analog sample. The design focuses on an analog hardware implementation and it is based on observing that the use of careful binary index assignment causes hyperplanes associated with the formation of the same bit at the MLP output to have similar spatial orientation. The results show that MLPs can be implement with less computational complexity than Kohonen self-organizing map (KSOM).

Keywords: MLP, video frame, vector quantization

I. INTRODUCTION

Image/video compression is always performed for the storage/transmission of data that has been already converted into the digital domain. Therefore the conventional data processing that stays between the sensor, usually a CMOS smart-pixel array, and the output compressed bit stream can be divided into two stages: A/D conversion, with as high resolution as possible [1], and signal compression. By implementing parts of signal compression operations such as matrix-vector multiplications, scalar additions and hard-limiting directly in the analog domain with accurate analog techniques currently available, one can eliminate the high-resolution A/D converters and part of the digital hardware/software for image/video compression. This leads to large savings in power consumption, circuit complexity and transmission bandwidth (or storage space). Recent papers proposes a multilayer perceptron (MLP) method to perform direct mapping from an analog image block into a vector index, implement vector quantization (VQ) without the need for carrying out first an analog-to-digital (A/D) conversion of each analog sample. The design focuses on an analog hardware implementation and is based on observing that the use of a careful binary index assignment causes hyperplanes associated with the formation of the same bit at the MLP output to have similar spatial orientation. Computer simulation results from a gray level video database will show that, for the same rate-distortion performance, MLPs can be implemented with less computational complexity than Kohonen self-organizing map (KSOM) based approaches advanced earlier.

The Kohonen self-organizing map has been proposed earlier for simple VQ with good adaptivity properties in [2],[3] and [4]. Many other applications of neural networks to image/video compression can be found in literature, including learning VQ (LVQ) [5],[6]. The IC implementation of an analog VQ by Cauwenberghs and Pedroni [7] is essentially an analog CMOS implementation of an offline-trained KSOM for generic VQ. Also, in [8] an analog CMOS implementation of a multilayer perceptron is made by Lont and Guggenbuhl.

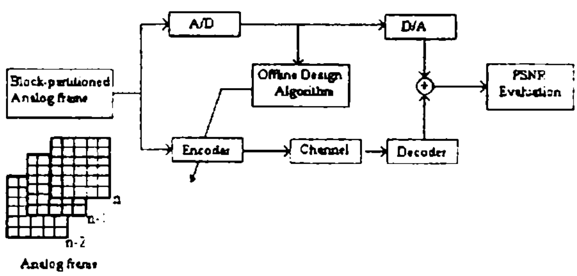

Figure 1: Direct Block Encoding of Analog Vectors

In this work, each image in the video stream is partitioned in *PxQ* blocks and the difference between current mean and estimated mean for each block is *Bo* bit scalar MLP encoded. Also, quantization of the analog residual vectors (obtained by subtracting the mean estimate from all pixel values of the block) is performed by an MLP which direct map the analog residual vectors into *B1* binary codeword. In this work we propose the use of a multilayer perceptron (MLP) with a simple, one hidden layer structure, for a very low complexity residual block VQ directly from the analog vector into binary codewords. Besides the elimination of the A/D bank, which is used only for offline design of the encoder (Fig. 1), the MLP binary outputs form the binary index of the nearest codevector directly by hyperplane decisions on the

[1] Dept of Telecommunications, Faculty of Electronics and Telecommunications, Technical University "Gh. Asachi" Iasi, B-dul Carol I, Nr. 11, Iasi 700506, ROMANIA
Email: fbeldianu@etc.tuiasi.ro

input data. This keeps the fully parallel encoding structure, eliminating the M hidden units involved in the KSOM winner-take-all (WTA) operation. These units are replaced by $\log_2 M$ hard-limiters at the network output layer, and the proposed MLP is computationally more efficient than the KSOM for the same distortion and amount of image compression.

II. PROPOSED IMPLEMENTATION

In this work, the 352×288 frame (CIF format) is partitioned into 4×4 (or 8×8) pixel blocks. Luminance blocks have 4×4 (or 8×8) pixels so 16 (or 64) is the number of pixels per block. The following operations are performed on a luminance block :

- find block mean $\mu_k(n)$ and difference from its previous estimate:

$$d\mu_k(n) = \mu_k(n) - \hat{\mu}_k(n) \qquad (1)$$

- *Bo* bit scalar quantization of the difference of the means $d\mu_k(n)$;
- subtracting the mean estimate from all pixel values of the block:

$$dI_k(n) = I_k(n) - \hat{\mu}_k(n) \qquad (2)$$

- *B1* bit vector quantization of the difference $dI_k(n)$ between the vector of pixel values and the mean estimate;
- upda.ing .he mean estimate:

$$\begin{aligned} \Delta\mu_k(n) &= \Delta\mu_k(n-1) + d\mu_k(n) \\ \hat{\mu}_k(n+1) &= \mu_k(n) + \Delta\mu_k(n) \end{aligned} \qquad (3)$$

Index k indicates the k-th block in the frame, n-1 indicates previous frame, n current frame and n+1 next frame.

III. MLP DESIGN

In order to avoid the complexity of the WTA operation and eliminate the demultiplexing block required for mapping the i-th winner index into a $\log_2 M$ binary codeword, a preliminary idea is to train an MLP to directly generate the binary word given an analog input. In the next two paragraphs two MLP configurations are presented, first is used to encode 4×4 residual pixel blocks $dI_k(n)$ and difference of the means $d\mu_k(n)$ and the second one is used to encode 8×8 residual pixel blocks. The proposed model for the estimation of block mean reduce the dispersion of the residual error, and so, improve the compression using a reduce number of 3 to 4 bits for quantization of the mean difference.

A fast design can be done by calculating the KLT of the given training data-set, and designing the weights of the input layer of the MLP explicitly as the components of the principal eigenvectors of the expansion design [9]. For VQ of high dimensional data according to a nearest-neighbor scheme, the use of a suitable index assignment such as pseudo-Gray encoding ([10],[11]) tends to cause all the hyperplanes associated with the formation of the same bit at the neural network output to have the same spatial orientation. Therefore, considerable simplification of the MLP structure can be systematically achieved at the expense of some additional distortion by replacing hyperplanes that have mainly the same direction by a single hyperplane or a reduced set of them.

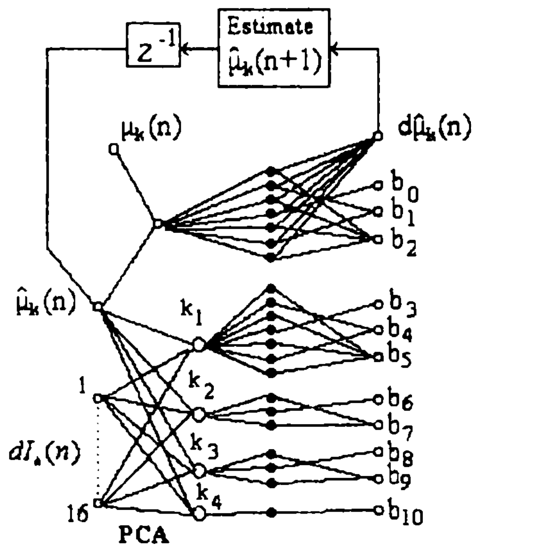

Figure 2: Design a MLP for analog VQ of residual 4×4 pixel blocks (first configuration). Wheights in the first layer are eigenvectors of the KLT expansion and in the second layer have to be design according to the Lagrangian optimization. White circles are linear and black circles perform thresholding.

A. *Encoding of 4×4 blocks*

In first configuration an MLP (Fig. 2) is designed to encode encode 4×4 residual pixel blocks $dI_k(n)$ and difference of the means $d\mu_k(n)$. First 4 components of the KLT of the 16-D residuals are preserved, for which 8 (3+2+2+1) bits are non-uniformly allocated, according to the well-known Lagrangian optimization method and 3 bits are allocated for the difference of the means. For this configuration , the computational complexity corresponds to the 17×4 multiplications required for PCA at the input layer, 1×7 multiplications for scalar encoding of $d\mu_k(n)$ and 14 (1×7+1×3+1×3+1×1) multiplications for scalar encoding of each principal component, so that the total number of multiplications is 89 for each block.

The total number of bits per frame is (352/4)×(288/4)×11=69696 bits/frame, which gives an encoding bit-rate up to 6366×25=1742.4 Kbps.

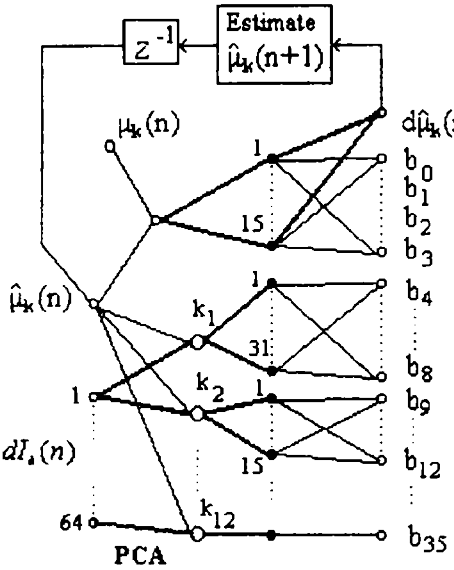

Figure 3: MLP for analog VQ of residual 8×8 pixel blocks (second configuration). Only parameters in bold line have to be design. The others are structurally given. White circles are linear nodes and black circles perform thresholding.

B. *Encoding of 8x8 blocks*

In the second configuration an MLP (Fig. 3) is designed to encode 8×8 residual pixel blocks. First 12 components of the KLT of the 64-D residuals are preserved, for which 32 bits are non-uniformly allocated and 4 bits are allocated for the difference of the means. The computational complexity is 65×12 multiplications required for PCA at the input layer, 1×15 multiplications for scalar encoding of $d\mu_k(n)$ and 96 (31+2×15+3×7+4×3+2) multiplications for scalar encoding of each principal component, so that the total number of multiplications is 876 for each block at step *k*.

For the second configuration results are (352/8)×(288/8)×36=57024 bits/frame, which gives an encoding bit-rate up to 57024×25=1425.6 Kbps.

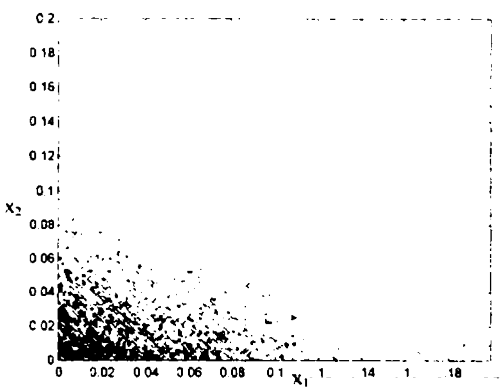

Figure 4a: Plot of the projections on first two principal components [x_1 x_2]

V. EXPERIMENTAL RESULTS

For computer experiment we used several video sequences, each video sequence having at least 25 frames. After computing residual vectors and mean differences, eigenvectors of the KLT expansion are found. In Fig. 4a the projections on first two principal components are plotted and there is evident the Laplacian distribution of them (Fig. 4b). In Fig. 5 quantization of the projections on two principal components is made by partitioning 2-D input space [x_1 x_2] according to their distribution, which is Laplacian with approximation.

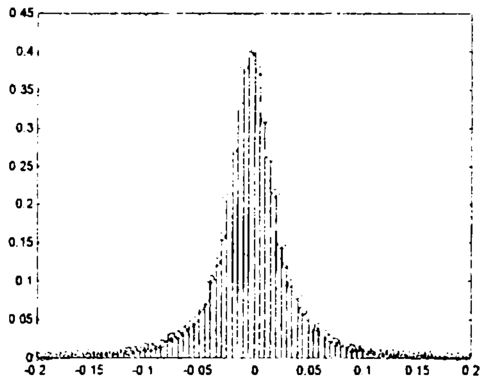

Figure 4b: Marginal distribution of x_1

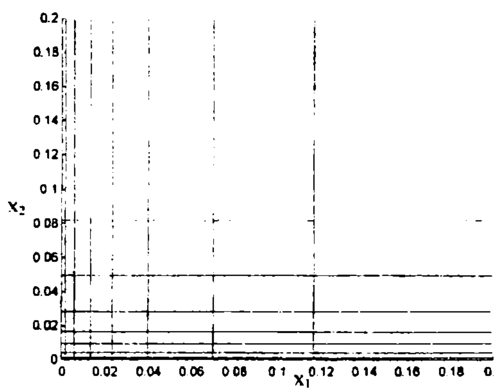

Figure 5: Partitioning of a 2-D input space [x_1 x_2]

In Table 1 MLP configuration I achieves a rate distortion (PSNR) of 29.34 dB and rate of 1742.4 Kbps and MLP configuration II achieves a rate distortion of 31.90 dB and a rate of 1425.6 Kbps. First configuration has slightly less than half of computational complexity of the second configuration.

Table 1

Configuration	Rate	PSNR	Multiplications per pixel
MLP 4×4	1742.4Kbps	28.34	5.56
MLP 8×8	1425.6 Kbps	31.90	13.59

In [8] a KSOM is trained to encode pixel blocks with a rate distortion (PSNR) of 31.44 dB and 31.56 multiplications per pixel.

So MLPs allow a very flexible exploitation of different rate-distortion and complexity trade-offs, at the same time keeping the overall complexity well below that of self-organizing systems.

We note that all bit-rates shown in Table 1 are calculated for source-encoded data only, and do not account for any entropy coding in either scheme, so that we can expect additional performance improvement by either off-sensor entropy coding or, more interestingly, embedding some primitive entropy coding in the MLP design.

V. MAIN CONCLUSIONS AND FUTURE WORK

In this paper we proposed a vector quantization scheme amenable to analog hardware implementation, namely a multilayer perceptron with one hidden layer allowing for direct mapping of an analog sample vector into a channel binary codeword (index). The numerical results show that MLPs can achieve the same rate-distortion performance of KSOM while reducing the number of multiplications per pixel. This advantage comes from the fact that the MLP partitions the data space using hyperplanes, and a careful assignment of index to the codebook vectors allows for a very efficient partition of data space using hyperplanes that are only approximately parallel along the principal component they relate to, i.e. that have roughly the same direction, which can be changed for performance optimization.

For future work, we plan to make the system adaptive, with a structured codebook that can be adjusted for different input image characteristics. Also bit rate can be reduced by using a lossless compression applied at the output of the MLP. A more sophisticated approach is to approximate the motion of the whole scene and the objects of a video sequence by using some basic motion compensation and motion estimation techniques. This gives much better residuals than a simple subtraction. So the residuals can be encoded at a lower bit-rate with the same quality or at the same bit-rate with a better quality. Also we plan to address the use of more efficient analysis systems such as Kernel PCA, which will replace linear PCA [12].

REFERENCES

[1] LM9628 Color CMOS Image Sensor VGA 30 FPS, *National Semiconductor*, May 2002

[2] S. Haykin, *Neural Networks, a comprehensive foundation, second edition*, Prentice Hall, 1999

[3] N. M. Nasrabadi and Y. Feng, "Vector quantization of images based upon Kohonen self-organizing feature maps", in *Proc IEEE ICNN*, 1988, pp. 1101–1108.

[4] K. K. Truong and R. M. Mersereau, "Multilayer Kohonen image codebooks with a logarithmic search complexity", in *Proc. ICASSP*, pp. 2789–2792, 1991.

[5] S. Ahalt, A. Krishnamurthy, P. Chen, and D. Melton, "Competitive learning algorithms for vector quantization", *Neural Networks*, vol. 3, pp 277–290, 1990.

[6] N. Vlajic and H. C. Card, "Vector quantization of images using modified adaptive resonance algorithm for hierarchical clustering", *IEEE Transactions on Neural Networks*, vol. 12, no. 5, pp 1147–1162, September 2001.

[7] G. Cauwenberghs and V. Pedroni, "A low-power CMOS analog vector quantizer", *IEEE Journal of Solid-State Circuits*, vol. 32, no. 8, pp. 1278–1283, August 1997.

[8] J.B. Lont, W Guggenbuhl, "Analog CMOS implementation of a multilayer perceptron with nonlinear synapses", *IEEE Transactions on Neural Networks*, vol. 3, no. 3, pp. 457-465, May 1992

[9] J. Gabriel, R.C Gomes, S.K. Mitra, "Analog multilayer perceptron implementation of low complexity VQ for image compression", *International Conference on Image Processing*, vol.3, pp II- 279-82, Proceedings 2003.

[10] K. Zeger and A. Gersho, "Pseudo-Gray encoding", *IEEETrans. on Comm.*, vol 38, no. 12, pp. 2147–2158, Dec 1990

[11] C. Gentile and M. Sznaier, "Hyperborders in the ronoidiagram-based neural net for pattern classification," in *Proc. Int. Joint Conf. on Neural Nets.*, 2002, vol. 3, pp. 2231–2236.

[12] V. N. Vapnik, *Statistical Learning Theory*, Ed Wiley-Interscience, 1998