AN ANALOG ASSOCIATIVE MEMORY CHIP

FOR VQ IMAGE COMPRESSION

Loris Navoni - Monica Besana -Pier Luigi Rolandi
Innovative Systems Design Group-Central R&D - STMicroelectronics
Agrate Brianza (Milan), ITALY
{loris.navoni , monica.besana, pierluigi.rolandi}@st.com

ABSTRACT

This paper presents a hardware implementation of Full-Search Vector Quantization Image Compression using an associative memory chip based on analog flash technology.
Taking advantage of the features of this architecture, that performs a parallel search on 4K 64-elements codebook in 4.6 μs, encouraging results have been obtained in terms of perceived image quality and computation speed.

1   INTRODUCTION

Computation complexity is one of the major drawbacks of VQ image compression. In fact, one of the limits of many VQ techniques is their inadequate capability to process great amount of data in a way that can assure a good image quality level at an acceptable computation speed. Large dimension codebooks can improve the quality of the decoded image, but, on the other hand, the codebook searches complexity increases drastically.

An interesting solution to reduce this high VQ cost can be the use of an Associative Memory [1], a data storage device with inherent search capabilities since it is accessed by the contents of the memory itself. This property, combined with a parallel search characteristic, makes associative memories very suitable for a direct VQ hardware realization.

Efficient VQ implementations on associative memory have been already proposed [2]-[3]. An innovative solution, which adopts massively parallel analog computation realized by the Analog Associative Memory Chip (AAMC) [4], is presented. This device, based on the analog Flash technology, has been developed as a general purpose pattern matching system and successfully applied on Character and Word Recognition [5].

High storage capacity, customizable pattern size, full parallelism in computation and high performance speed make the AAMC very suitable for many applications regarding signal processing field, and implementations of VQ methods for Image Compression can take largely advantages of the features of this chip.

To evaluate the quality of the AAMC coded image a perceptual measure, based on the Human Visual System, has been adopted rather than other image quality indexes.

2   VQ AND ASSOCIATIVE MEMORY

In its basic definition [6], VQ compares each *vector* obtained breaking data to be encoded into small blocks, to a set of representative vectors called *codebook*. The address of the codebook entry most similar to the vector to be coded (i.e. minimizing a given distortion measure) is transmitted to the VQ decoder. The decoder, having a codebook identical to the encoder, reconstructs the coded vector using the received address for a simple access to a lookup table.

An efficient VQ implementation can be obtained using an Associative Memory [1], a neural network able to operate on a distortion measure minimization to find the best approximation of the given input (i.e. the same criterion used in VQ).

Considerable Associative Memory properties are:
- large storage capacity;
- nearest neighbor approach, which implies the capability to obtain approximated matches;
- parallel search;
- pre-defined learning phase.

Several hardware solutions have been designed, following both digital and analog approach. This paper proposes a Flash-based architecture [7] that perform high-resolution patterns (5 bit) analog computation.

3   THE ANALOG ASSOCIATIVE MEMORY CHIP

Considerations about the choice of analog computation for pattern matching problems are presented in [8] where the key points highlighted below are described:
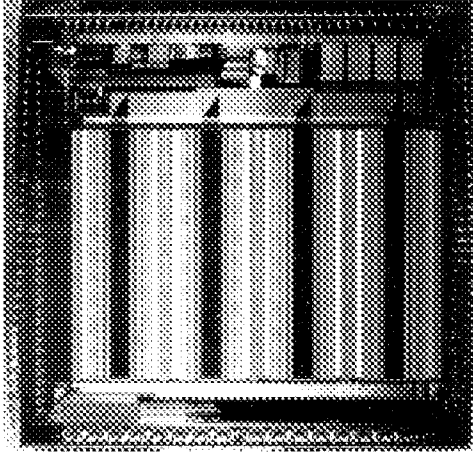
Fig.1: The Analog Associative Memory Chip

- analog computation is the choice on applications for which high efficiency is needed; on the other hand, this imply loss of generality of the design;
- 2-dimensional computing array is the most efficient architecture for analog computing;
- analog computing array imply the use of output summing;
- integration of both storage and computation into the same device is the most efficient approach to analog computation.

The AAMC [4], based on analog flash technology, contains 256K computing nodes, organized into 4K rows (vectors) each of 64 synapses. It is able to process in parallel the best match based on Manhattan distance between a 64-dimensional input vector and all 4K stored reference vectors. Each computing node is capable of storing and matching 5-bit data.

The device has been accomplished in 0.7 μm CMOS technology and it takes about 4.6 μs (55GCPS) to perform the computation with power consumption, at maximum speed, of 195 mW. A digital controller can enable or disable columns and lines, giving the possibility to make a selected computation.

Circuit behavior is explained in detail in [4] and [7], and can be summarized as follow.

A matching between input pattern and stored database is performed in parallel, based on minimum Manhattan distance (applied as comparison of Flash threshold voltages). Then a list of ordered indexes of the nearest elements of the input is returned. Optionally, subsequent criteria (i.e. clustering) can be adopted to compensate errors and confirm the best match choice.

The AAMC has been revealed as a good hardware solution for any kind of signal processing technique that needs a high degree of parallelism. A class of pattern matching problems that require to process a large amount of data in real time is an application field in which this chip can be used obtaining high performances.

4 THE ANALOG ASSOCIATIVE MEMORY CHIP AS VECTOR QUANTIZER

Main goal of the described project was to take advantage of the features owned by the AAMC to perform image compression using VQ.

Interest has been focused on Full Search VQ, due above all to the fact that others Vectors Quantization techniques needs great effort in pre and post processing, that was out of the scope of this work [9].

In the next sections a detailed description of the AAMC Vector Quantizer is presented

4.1 Prequantization

Gray level images are represented by a set of pixels, each usually codified with an 8-bit value. This resolution enables people to see images without any loss of information.

However, practical experience based on visual perception [10] suggest that an image can be well represented using less than 8 bit for each pixel and experimental results confirm that a reduction of the pixel resolution to 6-5 bits causes only a very small degree of perceived distortion (Figs. 2-3). This implies that the AAMC 5 bits resolution constraint is well suitable to image compression purpose. Therefore, a uniform scalar quantization to 5 bits is applied to the image data before coding.

4.2 Universal Codebook Generation

Associative Memory requires a learning phase where information are stored onto the computing nodes. Being Associative Memory an implementation of VQ, this phase is accomplished by a Codebook generator that creates values to be stored in the computing array.

A 4K-vectors codebook generation, each element representing a block of 4x4 of a generic gray-level image stored at 5-bit precision, has been performed using a Self-Organizing Maps algorithm [11], which distributes input vectors on a high dimensional space following statistical laws.

The choice of this method has been leaded by practical considerations since this work is focused on AAMC performance evaluations. Therefore, a small effort has been dedicated to codebook creation.

Fig.2: Original *Lena* 8-bit resolution

Fig.3: Scalar quantized *Lena* 5-bit resolution

4.3 Image Coding

During the coding phase, 5-bit gray level input image is divided in 4x4 blocks. Each block is "unrolled" on a 16-elements vector structure and then sent as input to the AAMC. There, it is compared with the stored codebook respect to the Manhattan distance and the index of the nearest vector is finally returned. In this way, the image is coded as a set of indexes.

Image reconstruction is simply obtained recovering the coded indexes from the loaded codebook.

The AAMC is able to store 4K vectors, 64 elements at 5-bit resolution each, which can be matched in parallel with a given pattern at speed of 217 K samples/sec. Due to the chosen 4x4-pixel image sampling, only 16 elements per line are used.

The achieved compression rate, which is obviously strictly dependent on the 4K vectors codebook size, is 0.75 bpp.

5 RESULTS

Image quality measures based on Human Visual System (HVS) take care of perceptual redundancies and irrelevancies, which are not usually considered by traditional image quality indexes.

To evaluate the AAMC performances an HVS-based method [12], that follows the *Just Noticeable Distortion* criterion [13], has been used. This index, named *Peak Signal-to-Perceptible-Noise Ratio* (PSPNR), exploits the human eye contrast sensibility and the texture masking effects and has been also useful to evaluate the effectiveness of the 5-bit image prequantization. Several test images have been compressed through the AAMC. Some results, reported in Table 1, show the good compromise between image quality and computation speed that can be guaranteed by the

described device, performing 512x512 pixels (5-bit per pixel) image compression in 75.3 ms. Fig. 4 presents an example of encoded image.

Considering the AAMC vector encoding rate of 217.3 Kvectors/s, speed performances for gray-level and true-color VGA and CIF format are described in Table 2.

Image	size	PSPNR
Pepper	512x512	27.77
Lena	512x512	27.26
Baboon	512x512	21.21

Tab.1: PSPNR values from AAMC simulations on well-known test images at 5 bit per pixel with 4K-vectors codebook.

Format	Size	Frame/s (gray-level)	Frame/s (true-color)
VGA	640x480	11.3	3.7
CIF	352x288	34.4	11.4

Tab.2: AAMC speed performances

6 CONCLUSIONS

A hardware implementation of the Full Search VQ Image Compression is described in this paper.

Despite the efficiency of VQ can be improved adopting different approaches (for example applying more sophisticated VQ methods), the use of an high parallel computational device, as the presented AAMC, should be considered a key point to reduce high VQ codebook search costs.

Interesting results of the AACM (its characteristics are summarized in Tab. 3) can drive further research on Still Image and Video application, also because VLSI

technology evolution should increase chip performances at very competitive levels

Chip Size	14x14 mm^2
Technology	0.7μm Flash CMOS
Power Supply	5 V
Power dissipation	195 mW
Codebook size	4096 codewords
Vector dimension	16 elements of 5 bit
Vector rate	217.3 Kvectors/s

Tab.3: Summary of Analog Associative Memory Chip characteristics

ACKNOWLEDGMENTS

7 REFERENCES

[1] J. M. Zurada, *Introduction to Artificial Neural Systems*, Chap. 6, pp. 313-320, West Publishing Company, 1992.

[2] S. Panchanathan-M. Goldberg, "A Content-Addressable Memory Architecture for Image Coding Using Vector Quantization", *IEEE Trans. Signal Processing*, Vol. 39, No. 9, pp. 2066-2078, September 1991.

[3] J. E. Fowler Jr. et al., "Real-Time Video Compression Using Differential Vector Quantization", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 5, No. 1, February 1995.

[4] A. Kramer et al., "55 GCPS CAM Using 5-bit Analog Flash", *ISCC97 Conference, Digest of Technical Papers*, pp. 44-45, Feb. 1997.

[5] L. Navoni et al., "Words Recognition using Associative Memory", *Proceedings of ICDAR97 Conference*, pp. 97-101, Aug. 1997.

[6] A. Gersho-R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publisher, 1992

[7] A. Kramer et al., "Ultra-Low-Power Analog Associative Memory Core Using Flash EEPROM Based Programmable Capacitors", *Proceeding of ISLPD 95 Symposium*, April 1995.

[8] A. Kramer, "Array-Based Analog Computation: Principles, Advantages and Limitations", *Proceedings of MicroNeuro '96*, February 1996.

[9] N. .M. Nasrabadi - R. A. King, "Image Coding Using Vector Quantization: A Review", *IEEE Trans. on Commun.* , Vol. 36, No. 8, pp. 957-971, August 1988

[10] A. K. Jain, "Fundamentals of Digital Image Processing", Prentice-Hall International, pp. 119-123, 1989.

[11] T.Kohonen et al., SOM_PAK The Self-Organizing Map Program Package, v.3.1, April 1995.

Fig.4: Reconstructed *Lena* at 0.75 bpp from 5-bit resolution input image

[12] C. Chou - Y. Li, "A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile", *IEEE Trans. on Circuits Syst. Video Technol.*, Vol. 5, No. 6, pp. 467-476, December 1995.

[13] N. Jayant - J. Johnston - R. Safranek, "Signal Compression Based on Models of Human Perception", *Proceedings of the IEEE*, Vol. 81, No. 10, October 1993.