A Computational Approach to Multimedia Communication Networks

Moshe Porat
Vision and Image Sciences Laboratory,
Department of Electrical Engineering Technion,
Israel Institute of Technology,
Technion City, Haifa 32000, ISRAEL
mp@ee.technion.ac.il, mporat@ieee.org

***Abstract*: - Multimedia communication is becoming a common means of human interaction. This increasing amount of information transmitted over wireless and land-line communication channels has to be handled efficiently to avoid congestion and delays in the network service. In this work, a computational approach to multimedia data transmission is proposed. One of its basic assumptions is that in many situations the multimedia information could be decomposed into *basic components or building blocks that represent a significant part of the data. To illustrate the* capabilities of the proposed approach, a compression system for video streams is investigated and tested. The system is based on the observation that in most multimedia systems, events of recent history of the stream can be used to describe present and future details of the multimedia information. Experimental results indicate a high compression ratio of more than 100:1, obtaining almost lossless visual reproduction with less than 0.1% distortion. The proposed approach can be similarly used also for other components of the data stream. Our conclusion is that the proposed computational approach could be efficient in presently used communication systems.**

Key-Words: Multimedia, Data Compression, Modeling, Vector quantization, Internet, Network channels.

I. INTRODUCTION

Multimedia tools play an important role in visual information systems. In several cases of data coding, such as in speech coding and facsimile, it is known that better compression might be available if additional information related specifically to the source of the transmitted data is considered. Usually this was done by developing a model for the information source and coding the parameters of the model instead of the raw information. Such models as exist in the image processing area rely on general assumptions like spatial high correlation, as used in the Markov process model [1]. There are other models related to the influence of blurring and defocusing, for example, but not many attempts have been made to model image sources, except perhaps a few works related to the human face as consists of several moving objects [2]. These models are of limited use. The basic assumptions, even though stated as general as possible, do not fit into many images, and by their very nature do not rely on information related directly to specific applications for which the visual communications system is designed.

A review of more sophisticated approaches to image coding is "Second Generation Image-Coding Techniques" by Kunt et al. [3], however, with regard to still images. One of the emphases in this review is on coding edges as an important feature of the visual information. A similar approach to coding of oriented edges is reported also by Giunta et al. [4, 5] for the case of low bit rate coding. Those methods are based on decomposition of the sequence into several bands [6], with various techniques applied to different bands. This approach of subband decomposition is in accordance with many findings related to the basic structure of the human visual system, where cells and groups of cells are sensitive to limited spatial-frequency bands, and are likely to be parts of different processing mechanisms of the human pattern recognition systems.

Accordingly, a need exists for a multimedia coding system which encodes and communicates at a lower bit rate and uses information about the specific characteristics of the source, the viewer, and the visual system, to reproduce the information with minimal distortion. In this paper a new approach is introduced based on the above principles and motivation.

II. THE MODEL AND THE SYSTEM

The model used in this study is based on specific properties related to the nature of multimedia streams. In particular, the main assumption is that changes or events of recent history are likely to repeatedly appear in the same area of the image where they have previously appeared. If not similarly repeated (up to allowed distortion), they are likely to be encoded as transformed versions of previous details [7]. Furthermore, if a specific event cannot be matched accurately according to coarse partitioning of the frame, a more refined description can be used [8]. According to this model a frame is divided into blocks and sub-blocks as shown in Figure 1.

In this figure an image is divided into sub-blocks or vectors of 4x4x4 pixels. In this figure, sixteen vectors comprise a block.

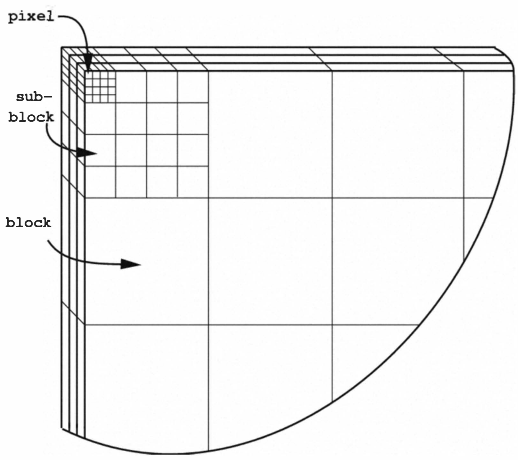

Figure 1. Blocks and sub-blocks. Each block and sub-block contains 4 consecutive frames.

Typically, an image is divided into at least 25 or 36 blocks, each plays a localized role in the compression of the information. Based on the above assumptions and structure, the proposed system is presented in Figures 2 and 3. In the first stage of the process (Figure 2) the blocks of the image after Image Partition are used for training localized codebooks (CB) [9]. These codebooks are based on the recent localized history of the sequence, thus contain significant information which can be readily used for quantizing the next frames of the sequence. The Delay (D) is needed to allow updating of the codebooks before quantization (Q). An illustration which indicates the size of the localized codebooks is shown in Figure 4. This example relates to a sequence of head-and-shoulder (Miss America) which represents a typical videophone interaction. In this illustration, codebooks of the background are small in size (indicated by dark gray and black) while codebooks of active areas like the head, and in particular the eyes and the mouth, are larger (light gray and white).

Most of the sub-blocks (vectors) of each block are adequately encoded in this stage by relatively sparse codebooks, with up to 512 code-words in the illustrated example. Some of the vectors, however, as shown in Figure 5, require additional attention due to distortion above a pre-determined threshold. These vectors are shown in the example of Figure 5 by black areas. Usually most of these blocks can be encoded using adjacent codebooks as shown schematically in Figure 3.

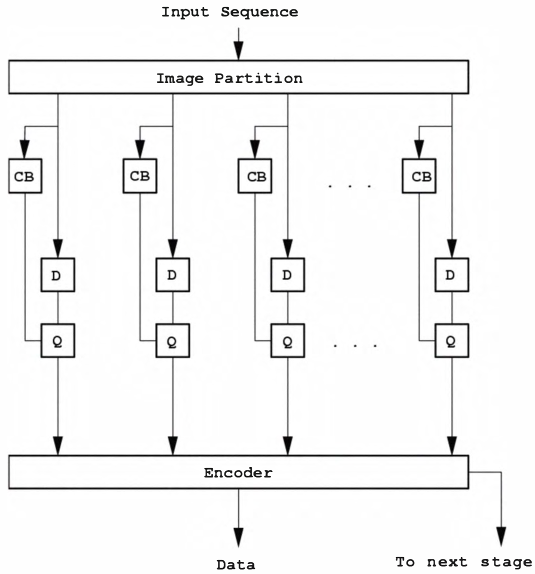

Figure 2. Block diagram of the first stage of the system.

Localized codebooks (CB) are trained by the recent history of the sequence, and then used to encode the next frames. D indicates Delay, and Q represents Quantizer.

It is assumed that these vectors refer to motions which have crossed the borders of their block (codebook) thus can be found in one of the adjacent codebooks, possibly after Transformation (T) of rotation or scaling.

The model has several parameters. In terms of history, two parameters relate to the duration of the history used, and to the typical duration of each movement, respectively. The latter also determines the minimum expected delay of the system. A localization parameter relates to the size of areas considered as likely to contain repeated motions during the time period defined as history. In the next stages of the system, a refined process is carried out with regard to those areas which were not encoded adequately in the first two stages.

Usually only very small number of vectors require this additional process. A pyramidal approach that can be adopted here may introduce additional parameters, mainly the factor by which the size of the blocks is changed between adjacent levels of the pyramidal representation.

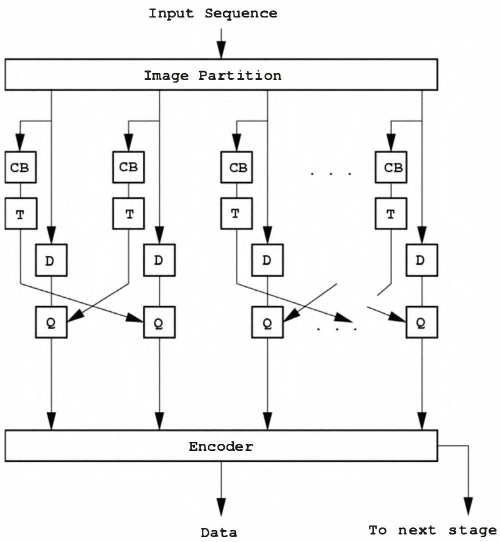

Figure 3

In the second stage of the system, adjacent localized codebooks are searched for best match, using – if required – Transformed (T) versions of codewords. CB, D and Q are as indicated in the first stage.

III. IMPLEMENTATION DETAILS AND RESULTS

The system was implemented according to the structure illustrated in Figures 2 and 3 with three pyramidal levels. The length of the history used for training the codebooks was of one second, with delay of 4 frames (about 1/10 of a second).

To avoid the need to transmit the codebooks to the receiving end, the recent history of the transmitted sequence was used for training the codebooks so that the same codebooks could be created at the remote receiver, thus only an index representing each codeword was transmitted via the transmission line [10]. Further reduction in the bit-rate is obtained when the index numbers of the code-words is encoded using the Huffman entropy-based approach.

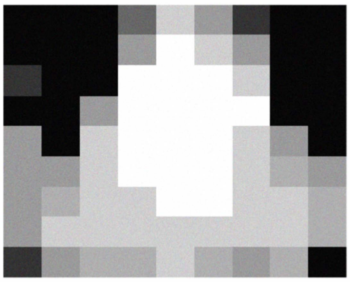

Figure 4. Example of the size of the codebooks used in the basic level of the encoder.

There are 9x9 blocks in this example (head-and-shoulder scene of Miss America, see next figures), where the grey level of each block represents the number of code-words used. The dark areas use small size codebooks, brighter areas use larger codebooks. The maximum number of codewords in a codebook here is 512 (white).

Figure 5

Reconstructed frame based on vectors found in localized codebooks, as shown in the block diagram of Fig. 2. Black areas were not encoded properly in this stage and are searched for best match in adjacent codebooks in the second stage (Fig. 3).

Typical results are shown in Figures 6-8. To afford comparison to previously proposed schemes, a three-dimensional vector quantization with one global codebook is used in the example of Figure 6. In this example, small vectors of 4x4x4 pixels were used obtaining high quality results at a compression ratio of 62:1. In comparison, the new model-based approach is presented in Figure 7. This localized approach provides a much higher compression ratio of more than 100:1, with the same quality of the perceived information. A graph of the algorithm's performance is shown in Figure 8.

Figure 6

Original (left) and reconstructed (right) frame from the sequence Miss America, using a global codebook with 4x4x4 blocks. The compression here is 62:1.

Figure 7

Original (left) and reconstructed (right) frame from the sequence Miss America, using localized codebooks with 4x4x4 blocks. The compression is 100:1.

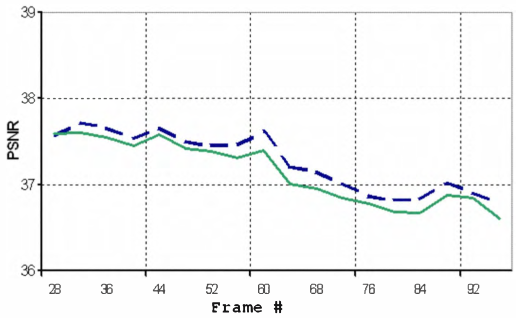

Figure 8

PSNR of the proposed algorithm as a function of the frame number. The basic algorithm is in dashed line. Solid line represents the results when an entropy-based (Huffman) technique is used.

IV. SUMMARY AND CONCLUSIONS

This paper has presented a computational approach to data coding using the history of the information sequence as building blocks for future encoding. In addition to the quality of the resultant transmission, the implementation of this approach can be systematically organized in a parallel manner by its very nature since different codebooks are created for each block and searched for independently. This computational approach is well supported by present days VLSI and network architecture, which allows systematical implementation of parallel systems. It should be noted that even for communication by serial machines, reduced complexity is achieved by processing of many small codebooks instead of a combined one [11]. Based on its performance, it is suggested that the new localized computational approach to data coding be further analyzed and integrated into presently available methods and networks [12].

ACKNOWLEDGMENTS

This research was supported in part by the Ollendorff Minerva Center. Minerva is funded through the BMBF.

REFERENCES

[1] Y.-Q. Shi and H. Sun, Image and video compression for multimedia engineering, CRC Press, 2000.

[2] A. Murat Bagcia, Yasemin Yardimcib and A. Enis Cetin, Moving object detection using adaptive subband decomposition and fractional lower-order statistics in video sequences, Signal Processing, Vol. 82, pp. 1941-1947, 2002.

[3] Kunt, M., and Kocher, M., Second-Generation Image-Coding Techniques. Proc. of IEEE, 73 (4), 549-573, 1985.

[4] T. Sikora, Trends and perspectives in image and video coding, Proceedings of the IEEE, 2005.

[5] Giunta, G.T.R., Image Sequence Coding using Oriented Edges, Image Communication 2 (4); 429-440, 1990.

[6] O. Gendler and M. Porat, "On Efficient Quantization for Image Recompression", the Eusipco proceedings, Glasgow, 2009.

[7] Porat, M., and Zeevi, Y.Y., The Generalized Gabor Scheme in Biological and Machine Vision. IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-10 (4), 452-468, 1988.

[8] Katzir, N., Lindenbaum, M., and Porat M., Curve Segmentation under Partial Occlusion. IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-16 (5), 513-519, 1994.

[9] Porat, M. and Zeevi, Y.Y., Localized Texture Processing in Vision: Analysis and Synthesis in the Gaborian Space. IEEE Trans. on Biomedical Engineering, BME-36 (1), 115-129, 1989.

[10] Linde, Y.L., Buzo, A., and Gray, R.M., An Algorithm for Vector Quantizer Design. IEEE-COM, 28, 84-95, 1980.

[11] Furman, D., and Porat, M., On Content-Based Very Low Bitrate Video Coding. The Very-Low Bitrate Video (VLBV) Conference, Madrid, 2003.

[12] Kirshner, H. and Porat, M., "On the Role of Exponential Splines in Image Interpolation", IEEE Trans. on Image Processing, vol. 18, no.10, pp. 2198-2208, 2009.