論文2003-40SP-1-9

A Self-Organizing Model Based Rate Control Algorithm for MPEG-4 Video Coding

장 지 밍 *, 장 승 기 *, 박 정 훈 *, 김 용 제 *

(ZhiMing Zhang, Seung-Gi Chang, Jeong-Hoon Park, and Yong-Je Kim)

요 약

본 논문에서는 자기구성 뉴로퍼지 네트워크를 이용한 MPEG-4 비트율 제어알고리즘을 제안한다. 경험적인 수식을 바탕으로 rate-distortion(RD) 모델을 구성하는 일반적인 방법과는 달리 제안하는 알고리즘의 기본적인 아이디어는 온라인으로 RD모델을 스스로 구성하고 매 프레임마다 그 구조를 적응적으로 업데이트 하는SOLPN을 이용해 RD 모델을 구현하는 것으로 많은 비트율 제어 방식 중 프레임을 기반으로 한 비트율 제어만을 본 논문에서는 고려한다. 특히 이 알고리즘은 오프라인에서 미리 트레이닝하는 것이 필요가 없기 때문에 실시간 코딩에도 적용 가능하다. VM18.0과의 비교 실험 결과들을 보면 본 논문에서 제안하는 비트율제어 알고리즘이VM18.0[16]에 비해 주관적인 화질 향상뿐만 아니라 적은 프레임 스킵(frame skip)과 높은 PSNR을 나타낸다.

Abstract

A new self-organizing neuro-fuzzy network based rate control algorithm for MPEG-4 video encoder is proposed in this paper. Contrary to the traditional methods that construct the rate-distorion(RD) model based on experimental equations, the proposed method effectively exploits the non-stationary property of the video date with neuro-fuzzy network that self-organizes the RD model online and adaptively updates the structure. The method needs not require off-line pre-training; hence it is geared toward real-time coding. The comparative results through the experiments suggest that our proposed rate control scheme encodes the video sequences with less frame skip, providing good temporal quality and higher PSNR, compared to VM18.0.

Keyword : self-organizing neural network, MPEG, rate control

Ⅰ. INTRODUCTION

In real-time video communication, the rate control strategy has become a key and plays a critical role in the video encoder. In order to predict the output, it is very desirable to be able to construct a source model that can estimate the bits produced by a video encoder for a chosen set of coding parameters.

Several RD models for DCT-based video encoders have been proposed in the literature. Some of them (Ref. [6][10]) encode each image block several times, and then intelligently select the best parameters, but these approaches are generally not suitable for real-time encoding because of their high computational complexity. Other approaches (Ref. [1][3][4]) select the quantizers based on different mathematical models, where the control parameters are estimated from the observed RD data of the coding system. While they are adequate in real-time applications, they can suffer from frequent frame skipping and wasting of channel bandwidth in low-delay

* 정회원, 삼성전자 DM연구소 Connectivity Group.
(DM R&D center, Samsung Electronics)
接受日字:2002年5月13日, 수정완료일:2002年10月29日

applications. Also, it is noted that to achieve high coding performance, the RD algorithms are becoming increasingly complex and more and more experiments need to be done to get the experimental control parameters. In order to simplify the construction process, Ref. [2] proposed a self-organizing feature map based rate control scheme, which constructed the frame-based global RD model by a neural classifier. Although it obtains a good result, Ref. [2] requires off-line training, i.e. using multiple video samples with QP varying from 1 to 31 to collect possible variations in the image characteristics, which is not easy to update the configuration and corresponding neural network structure.

In this paper, we propose a simple, yet effective self-organizing neural network based rate control scheme for DCT-based video coders in real-time video communication environments. Our method focuses on the frame-based rate control and the basic idea is to self-organize the neural network that accurately predicts the target number of bits and encodes the series with less skipped frames. We implement our new rate control strategy in a MPEG-4 encoder, and compare its performance to that of the well-known rate control algorithm in VM18.0[9,11].

The paper is organized as follows: in Section II, we introduce the basic idea on self-organizing neural network. Section III develops a rate control algorithm with this kind of neural network. Simulations are carried out in Section IV. Finally, Section V is devoted to conclusions.

II. SELF-ORGANIZING NETWORK

The main reason behind our method is that there exist similarities among the adjacent frames, which can be well modeled as a fuzzy system. At the same time, the requirements that the inference system needs not only the modeling accuracy, but also the learning ability that can quickly and adaptively catch the non-stationary nature of the video lead us to a neural network-based solution.

Recently, bringing the learning abilities of neural networks to automate and realize the design of fuzzy logic systems has become a very active research area. However, current neuro-fuzzy systems have some limitations. First, the number of partitions of each input variable is determined *a priori*, which means that the number of rules is fixed. Second, the back-propagation for the current fuzzy model is usually very slow. Normally, the difficulties grow for the following reasons: First, the gradient descendent learning scheme adopted by the BP network is inherently a slow learning process. Second, the massive distribution of the BP network requires the repeated presence of the training samples in order to make the learning process convergent, leading to the difficulty in self-organizing the system from the data(Ref. [5][8]).

To avoid these problems, this paper adopts the method from Ref. [12], which describes an improved learning method for a self-organizing neuro-fuzzy system. The method is based on Kohonen's learning algorithms (Ref. [8]), Counter-Propagating Network (CPN, Ref. [5]) and back-propagation algorithm with adaptive learning rate (Ref.[7]). The adaptive gradient descent method has the merits of fast convergence rate and easy computation. By combining both the nearest-neighbor clustering scheme and the adaptive gradient descent method, the learning process converges much faster than the original back-propagation learning method. It is just because of its simplicity and accuracy that we use this method to construct the RD model and adaptively update the structure online.

III. RD MODEL

In typical video coding techniques, such as H.263, MPEG and other block-based video coders, the picture quality for each image block is strongly dependent on the block's quantization parameters and corresponding RD model.

For the frame-based rate control algorithm in MPEG-4 VM18.0, the quadratic RD model is used as follows (Ref.[9][11]):

$$T = MAD \cdot \left(\frac{X_1}{QP} + \frac{X_2}{QP^2} \right) \quad (1)$$

where T denotes the target bits, MAD is the sum of the absolute difference between the original image frame and motion compensated reconstructed image frame, QP is the quantizer, and X_1 and X_2 are RD modeling parameters that should be updated after finishing the encoding process for each image frame. Instead of using empirical equations in MPEG VM18.0, we just regard the video RD model as a time-varying nonlinear system mapping, which ca be modeled as a neuro-fuzzy system. Further, In order to construct an accurate frame-based RD model for discriminating image characteristics for specific video sequences, several new parameters in addition to T, MAD and QP are required. We found that there are

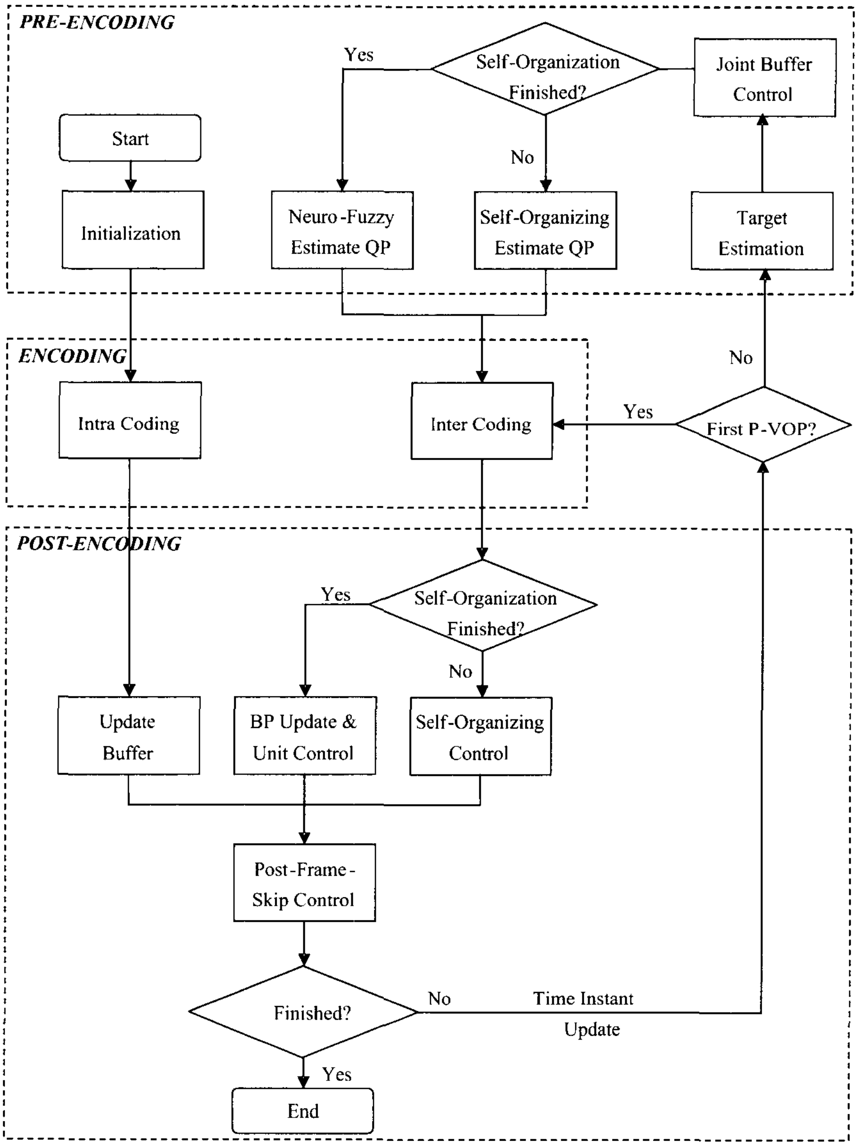


Fig. 1. Block diagrams of rate control algorithm
그림 1. 비트율 제어 알고리즘의 블록 다이어그램

strong relationships between the adjacent target bits allocation, so finally the global RD model for the inter-frame is modeled as follows:

$$PR_i = F(QP_{i-1}, MAD, T_i, T_{i-1}, R_{i-1}) \quad (2)$$

where F is the frame-based RD model to be approximated. T_i and T_{i-1} are target bits for the current frame i and previous frame $(i-1)$ respectively. R_{i-1} is the actually used bits for the last frame and PR_i is the bit-rate that is predicted by the system.

The structure of the encoder is shown in <Fig. 1.> (Only IPPP... series are considered here.)

The proposed rate control algorithm consists of following steps:

Step1 : Encode the first I and P frames using user-predefined quantizer.

Step2 : Update the RD buffer or post frame-skip control and collect the histogram data for the neural network.

Step3 : Estimate *QP* by self-organizing network or constructed neuro-fuzzy network:

1) Place the input vector to the input layer.
2) Get the prediction output from the neural network, and predict the target output with $QP \pm 2$ (In order to maintain a consistent view quality, we restrict the difference between the adjacent frames do not exceed ±2).
3) Find the winner: find the suitable *QP* that has the nearest output with target bit allocation.
4) Assign this *QP* as quantizer.

Step4 : Encode the current image frame.

Step5 : Update the RD model. If the encoded frames do not exceed *N* (a predefined constant for self-organizing phase), the neural network continues the self-organization. Else, the neural network is adaptively updated, and the total number of the neurons does not change anymore. Moreover, according to the output error, the structure or parameters of the constructed network will be changed as follows:

Because of the nature of the video sequence, to predict the output with enough accuracy is not easy, especially when there exists scene change. In order to solve this problem, the update scheme for the network is divided into two categories. If the prediction of neuro-fuzzy network is closer to the actual output, parameters in the network will be updated with online BP. However, if the error exceeds the threshold, current input record will overwrite the neural node that gets the largest similarity output. That means the fuzzy base structure will be changed while the size of the network remains the same. The objective of this kind of update scheme is to adaptively tune the network, which maintains the size of the network at a suitable level that avoids oversize, and also catches up the variation of the picture characteristics at the same time.

Step6 : Buffer control. If the buffer status is above 80%, the next frame will be skipped just as the standard VM18.0.

Step7 : Encode the next frame (go to step 3) or exit.

It should be noted that when encoding the first N frames, the model is quickly constructed with self-organization while the BP stage is not needed until the structure is stabilized. This kind of scheme avoids the complex computation and still maintains a good modeling accuracy.

IV. SIMULATION RESULTS

We used the standard MPEG4 codec with H.263-type quantization and tested: 1) the new self-organizing model based rate control, and 2) the rate control in VM18.0. Several QCIF video sequences (source frame rate is 30 f/s) that are widely used in MPEG applications were used in the simulation. The first frame was intra-coded with quantization step 10 and the other frames were all inter-coded. The test was done on following cases:

- ◆ Case 1: frame rate is 10f/s and target bitrate is 48kbits/s
- ◆ Case 2: frame rate is 15f/s and target bitrate is 64kbits/s
- ◆ Case 3: frame rate is 30f/s and target bitrate is 128kbits/s

Table 1 shows comparisons of the average PSNRs (in dB) of encoded bitstreams by the MPEG-4 VM18.0 rate control and the proposed rate control algorithms in various configurations.

Table 1. Test results on PSNR.
표 1. PSNR 실험결과

Video Sequences	Test Case	Skipped Frames		Average PSNR (Y)		
		MPEG 4 VM18.0	This Work	MPEG-4 VM18.0	This Work	PSNR Gain (dB)
CONTAINER	1	0	1	35.65	36.11	+0.49
	2	0	0	35.92	36.46	+0.54
	3	27	8	37.48	37.80	+0.32
FOREMAN	1	1	0	30.86	30.66	-0.20
	2	0	0	31.24	31.33	+0.09
HALL	1	0	0	37.14	37.84	+0.70
	2	0	0	37.33	38.20	+0.87
	3	8	2	38.84	39.24	+0.40
MAD	2	0	0	38.32	38.46	+0.14
	3	23	0	39.69	39.82	+0.13
NEWS	1	0	0	34.17	34.66	+0.49
	2	2	1	34.76	35.24	+0.48
	3	0	0	36.72	37.47	+0.75
SILENT	2	0	0	34.57	34.94	+0.37
	3	6	2	36.53	37.18	+0.65

From the <Table 1> we can easily see that the proposed rate control scheme has better results in average PSNR and has less skipped frames. At the same time, it is shown that although our method uses fewer parameters than the other neural network-based algorithm (In Ref. [2], 6 parameters are used for constructing the RD model, whereas 5 parameters in our method) it still gets the better results. Moreover our method needs not the complex and time-consuming pre-training, which is suitable for the real-time processing.

It is also noted that in most of the cases our method gets the better PSNR than the MPEG-4 VM18.0, while in other case (such as bitstream "FOREMAN" in test case 1), the proposed method gets the less PSNR. But in case of "FOREMAN", it's difficult to get a certain answer that which method is more suitable, because the psychological effect of the video quality is not clear at current stage, which somewhat depends the motion of the picture. Indeed there exist some trade-offs between the temporal continuity and spatial resolution, and the complexity increases more especially in the MPEG-4 where the frame rate can be changed. We will address the optimization of frame rate, bitrate and the continuity of visual effect in the future work.

V. CONCLUSION

In this paper, a self-organizing neural network-based rate control scheme has been proposed. Based on the methodology, a frame-based RD model is self-organized with histogram data and is used to predict the output of the target bits, which is totally different from the traditional methods that construct the RD model by experimental equation. In order to adaptively catch up the constantly-change video characteristics, except the normal BP algorithm for optimization, the algorithm for quickly updating the node is also taken effect especially when there's a scene change. This updating scheme shows good flexibility and efficacy in the experiments for real-time rate control. In comparison to VM18.0 rate control, our method skipped fewer frames, encoded the sequences with higher PSNR. In particular, since this rate control skipped fewer frames, the video sequences are encoded with better motion continuity, providing good subjective image quality.

REFERENCES

[1] B. Tao, B. W. Dickinson, H. A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding", IEEE Trans. Circuits Syst. Video Technol., Vol. 10, No. 1, pp. 147~157, 2000.

[2] G. H. Park, Y. J. Lee, "Rate control algorithm using SOFM-based neural classifier", Electronics letters, Vol. 36, No. 12, pp. 1041~1042, 2000.

[3] J. R. Corbera, S. M. Lei, "A frame-layer bit allocation for H.263+", IEEE Trans. Circuits Syst. Video Technol., Vol. 10, No. 7, pp. 1154~1158, 2000.

[4] J. R. Corbera, S. M. Lei, "Rate control in DCT video coding for low-delay communications", IEEE Trans. Circuits Syst. Video Technol., Vol. 9, No. 1, pp. 172~185, 1999.

[5] J. Nie, "Constructing fuzzy model by self organizing counterpropagation network", IEEE trans. Syst. Man Cyber., Vol. 25, No. 6, pp. 963~970, 1995.

[6] L. J. Lin, A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics", IEEE Trans. Circuits Syst. Video Technol., Vol. 8, No. 4, pp. 446~459, 1998.

[7] M. Smith, "Neural networks for statistical modeling", Van Nostrand Reinhold, New York, 1993.

[8] T. Kohonen et al., "Engineering applications of the self-organizing map", Proc. of IEEE, Vol. 84, No. 10, pp. 1358~1383, 1996.

[9] T. Chiang, Y. Q. Zhang, "A new rate control scheme using quadratic rate distortion model", IEEE Trans. Circuits Syst. Video Technol., Vol. 7, No. 1, pp. 246~250, 1997.

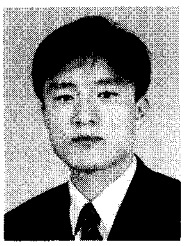
[10] W. Ding, B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling", IEEE Trans. Circuits Syst. Video Technol., Vol. 6, No. 1, pp. 12~19, 1996.

[11] W. P. Li, J. R. Ohm et al, "MPEG-4 Video Verification Model version 18.0", ISO/IEC JTC1/SC29/WG11/N3908, 2001.

[12] Z. M. Zhang et al., "An improved self-organizing CPN-based fuzzy system with adaptive back propagation algorithm", Fuzzy sets and systems, Vol. 130, No. 2, pp. 227~236, 2002.

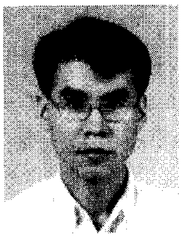
저 자 소 개

ZhiMing Zhang was born in Hubei province, China P.R. in 1972. He received B. S. degree in electronics and information engineering from Huazhong University of Science and Technology, China, in 1993, and the Ph. D. degree in electronic engineering from Beijing Institute of Technology, in 2000. In 2000, he joined Samsung Electronics, where he was engaged in video codec designs.



Seung-Gi Chang was born in Uljin, Korea in 1965. He received M.S and Ph.D. degrees in Electronics engineering from KyungHee University in 1995 and 2000, respectively. Since 2000, he has been a member of senior engineer with Digital Media R&D center, Samsung Electronics, working on video compression. His current research interests include video compression such as MPEG-4, H.264 and multimedia communications.

Mr. Park, Jeong-Hoon was born in TongYoung City, Korea in 1968. He received the B.S. and M.S. degrees from Han Yang University, Seoul, Republic of Korea in 1991 and 1994 respectively, in Electronic Engineering. He joined the multimedia R&D center of Samsung Electronics company, Korea in 1994. From 1997 to 1999, he had worked for ITU-T H.263v3 video coding standardization, especially contributing to the part on error resilient video coding. From 1998 to 1999, he had worked at EE Dept. of University of California, Los Angeles, as a visiting researcher to conduct research on error resilient video coding and ITU-T video coding standardization. He also participated in the standardization of the mobile multimedia telephony and streaming specification of codec working group under 3GPP, and 3GPP2 from 1999 to 2001. Since February 2001, he has been a research engineer at the digital media R&D center of Samsung Electronics. He now serves as co-chair of the interoperability working group in the MPEG-4 Industry Forum. He currently leads a project in developing the mobile multimedia streaming service and terminal with various MPEG-4 related technology in Samsung Electronics.



Yong-Je Kim was born in Gyungnam, Korea in 1962. He received B.S degree from Sogang University in 1985. Since 1984, he has been a member of principal engineer with Digital Media R&D center, Samsung Electronics.