VIDEO CODING BASED ON SEAMLESS COMBINATION OF MOTION COMPENSATION AND MATCHING PURSUITS

Ichiro Matsuda and Susumu Itoh
Dept. of Electrical Engineering, Science University of Tokyo,
2641 Yamazaki Noda-shi, Chiba Prefecture, 278-8510, JAPAN
Tel: +81 471 24 1501 (ex.3722); Fax: +81 471 24 9367
e-mail: matsuda@itohws01.ee.noda.sut.ac.jp

ABSTRACT

This paper proposes a low bit-rate video coding scheme based on combination of motion compensation and waveform coding. In order to realize appropriate bit-allocation to both motion compensation and waveform coding according to the structural properties of video signals, we employ matching pursuits which can freely control a coding rate of motion compensated frame differences. Moreover we develop a new iterative motion compensation method which can also control a bit-rate flexibly and can successively improve quality of reconstructed images, and combine it with the above matching pursuits seamlessly and adaptively. In other words, motion compensation and matching pursuits are carried out alternatively and repeatedly to attain higher coding performance within a limited coding rate. Simulation results indicate that the proposed coding scheme shows SNR gains of 0.3–1.0 dB over the H.263 standard and provides visually better reconstructed images with less blocking artifacts than it.

1 INTRODUCTION

Most of recent video coding systems are based on a hybrid coding method which consists of two operations: motion compensation (MC) and waveform coding. In this method, it is generally difficult to realize optimum bit-allocation to both operations since waveform coding, typically based on the DCT, is applied to residual signals which cannot be obtained until the operation of MC has been completed. As a result, percentage of information on motion vectors tends to increase especially at lower coding rates and coding performance deteriorates rapidly [1].

To cope with this problem, this paper proposes a novel video coding scheme based on combination of MC and waveform coding. In this scheme, a waveform coding algorithm called matching pursuits (MP) [2, 3] is applied to motion compensated frame differences. The MP algorithm represents a signal as a linear combination of basis-functions which are iteratively selected from a dictionary, therefore it can freely control a coding rate by varying the number of the basis-functions. We directly introduce this algorithm into the operation of MC and moreover develop a new iterative MC method in order to carry out rate control flexibly for MC as well as MP. In the coding scheme, MC and MP are executed alternatively and iteratively. In other words, both operations of MC and MP are tested in each iteration of the coding process and suitable one is adaptively selected. This iterative coding operation can appropriately allocate the number of bits to MC and MP under the condition of either rate control mode: constant bit rate (CBR) or variable bit rate (VBR).

2 MATCHING PURSUITS

The MP algorithm [4] represents a signal $F(\boldsymbol{t})$ as a linear combination of basis-functions which are selected from an overcomplete dictionary $\boldsymbol{\mathcal{D}} = \{\, g_{\boldsymbol{\gamma}}(\boldsymbol{t}) \,\}$. When the number of the basis-functions is limited to M, an approximation of $F(\boldsymbol{t})$ is given by:

$$P_M(\boldsymbol{t}) = \sum_{k=1}^{M} a_k \cdot g_{\boldsymbol{\gamma}_k}(\boldsymbol{t}-\boldsymbol{\tau}_k), \tag{1}$$

where $\boldsymbol{\tau}_k$ is a shift parameter of the basis-function $g_{\boldsymbol{\gamma}_k}(\boldsymbol{t})$, $\boldsymbol{\gamma}_k$ is an index corresponding to a particular basis-function in the dictionary $\boldsymbol{\mathcal{D}}$ and a_k is an expansion coefficient. In order to obtain a better approximation of $F(\boldsymbol{t})$, the parameters $\boldsymbol{\tau}_k$, $\boldsymbol{\gamma}_k$ and a_k $(k = 1, 2, \cdots, M)$ are in general determined successively so that the absolute value of the following inner product can have a maximum:

$$a_k = \left\langle e_{k-1}(\boldsymbol{t}),\, g_{\boldsymbol{\gamma}_k}(\boldsymbol{t}-\boldsymbol{\tau}_k) \right\rangle, \tag{2}$$

where $e_k(\boldsymbol{t})$ is a residual signal obtained after the k-th stage of the approximation, that is

$$e_k(\boldsymbol{t}) = F(\boldsymbol{t}) - P_k(\boldsymbol{t}) = e_{k-1}(\boldsymbol{t}) - a_k \cdot g_{\boldsymbol{\gamma}_k}(\boldsymbol{t}-\boldsymbol{\tau}_k). \tag{3}$$

If all of the basis-functions in the dictionary $\boldsymbol{\mathcal{D}}$ are normalized, i.e. $||g_{\boldsymbol{\gamma}}(\boldsymbol{t})||^2 = 1$, the residual signal $e_k(\boldsymbol{t})$ is orthogonal to the function $g_{\boldsymbol{\gamma}_k}(\boldsymbol{t}-\boldsymbol{\tau}_k)$. Therefore Eq. (3) leads to the following equation:

$$a_k^2 = ||e_{k-1}(\boldsymbol{t})||^2 - ||e_k(\boldsymbol{t})||^2. \tag{4}$$

It means that maximizing the absolute value of a_k is equivalent to finding the optimum set of parameters $\{\boldsymbol{\tau}_k, \gamma_k, a_k\}$ which maximizes the difference of SSE (Sum of Square Error) between the current and the next approximations, namely, $\Delta SSE = ||e_{k-1}(\boldsymbol{t})||^2 - ||e_k(\boldsymbol{t})||^2$.

Recently, Neff and Zakhor applied the MP algorithm to coding of motion compensated frame differences and demonstrated that their MP-based coding outperformed the conventional DCT-based coding at low coding rates [2]. In the MP-based coding, M sets of the three parameters $\{\boldsymbol{\tau}_k, \gamma_k, a_k\}$ are determined in the above-mentioned way, therefore one can freely control a coding rate by dynamically changing a value of M. In order to realize flexible rate control in MC as well as MP, in this paper we introduce the MP algorithm into the following new iterative MC method.

3 ITERATIVE MOTION COMPENSATION

In our coding scheme, both MC and MP are used as operations to decrease the energy of residual signals. In other words, MC is regarded as a kind of vector quantization (VQ) whose codebook consists of waveforms in a reference frame. This interpretation has an analogy with the relationship between MP and a multistage gain/shape VQ technique [4]. From this point of view, we propose an iterative MC method which improves quality of reproduced signals successively and gradually like the MP algorithm does.

Let $C_k(\boldsymbol{t})$ be a reproduced signal at the time just after the iterative MC method has been carried out k times. And it starts from $C_0(\boldsymbol{t})$ which is a reference image, namely, a reconstructed image in the previous frame. The MC method is executed for a square block $\boldsymbol{B}(\boldsymbol{s}_k, n_k)$ as illustrated in Figure 1 where $\boldsymbol{s}_k$ and n_k are parameters which specify a position and size of the block respectively and $\boldsymbol{v}_k$ is a motion vector detected in the block. In addition, $\boldsymbol{t}$ is not time, but a 2-D positional vector hereafter. Accordingly, the reproduced signal $C_k(\boldsymbol{t})$ is renewed as follows:

$$C_k(\boldsymbol{t}) = \begin{cases} C_0(\boldsymbol{t} - \boldsymbol{v}_k) & (\boldsymbol{t} \in \boldsymbol{B}(\boldsymbol{s}_k, n_k)) \\ C_{k-1}(\boldsymbol{t}) & (\text{otherwise}). \end{cases} \qquad (5)$$

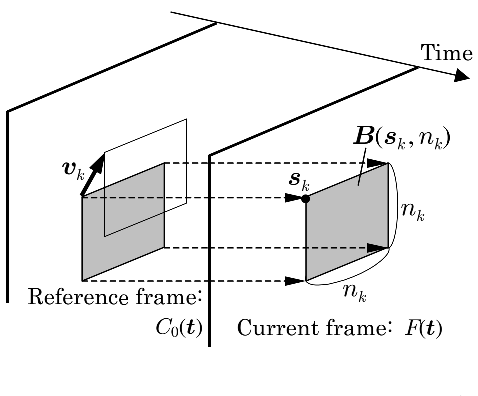

Figure 1: Iterative motion compensation method.

A set of the three parameters $\{\boldsymbol{s}_k, n_k, \boldsymbol{v}_k\}$ required for reproducing the signal $C_k(\boldsymbol{t})$ is determined so that a value of the following ΔSSE can have a maximum:

$$\begin{aligned} \Delta SSE &= ||F(\boldsymbol{t}) - C_{k-1}(\boldsymbol{t})||^2 - ||F(\boldsymbol{t}) - C_k(\boldsymbol{t})||^2 \\ &= \sum_{\boldsymbol{t} \in \boldsymbol{B}(\boldsymbol{s}_k, n_k)} \Big\{ \big(F(\boldsymbol{t}) - C_{k-1}(\boldsymbol{t})\big)^2 \\ &\qquad - \big(F(\boldsymbol{t}) - C_0(\boldsymbol{t} - \boldsymbol{v}_k)\big)^2 \Big\}, \end{aligned} \qquad (6)$$

where $F(\boldsymbol{t})$ represents the original image of the current frame. Since exhaustive search for the parameters $\boldsymbol{s}_k$, n_k and $\boldsymbol{v}_k$ which maximize ΔSSE needs an extremely large amount of computational load, we restrict search accuracy for the parameter $\boldsymbol{s}_k$ to every four pels and limit possible values of the parameter n_k only to 4, 8, 16 and 32 pels. In this case, a value of ΔSSE in a large block ($n_k \geq 8$) can be easily calculated by only adding four values of ΔSSE in its four subblocks, the size of which is $n_k/2 \times n_k/2$. Furthermore, search for the motion vector $\boldsymbol{v}_k$ in each block is needed only once, because the vector $\boldsymbol{v}_k$ which maximizes ΔSSE in Eq. (6) is determined independently of the latest reproduced signal $C_{k-1}(\boldsymbol{t})$. These facts contribute to remarkable reduction of computational load required for execution of the iterative MC method.

4 VIDEO CODING BASED ON COMBINATION OF MC AND MP

In this section, we describe a concrete procedure of our coding scheme. A block diagram of the encoder is shown in Figure 2.

4.1 Initialization of the Encoder

In the coding scheme, a reconstructed image obtained after the k-th stage of an iterative coding operation is given by the sum of two signals $P_k(\boldsymbol{t}) + C_k(\boldsymbol{t})$, where $P_k(\boldsymbol{t})$ and $C_k(\boldsymbol{t})$ are the components reproduced by MP and MC respectively. Before starting the iterative coding operation for each frame, $P_0(\boldsymbol{t})$ is initialized to zero in the whole frame, and the reference image which has been reconstructed in the previous frame is substituted into $C_0(\boldsymbol{t})$ as an initial image. Then a value k of a counter for the iterative operation is set to one.

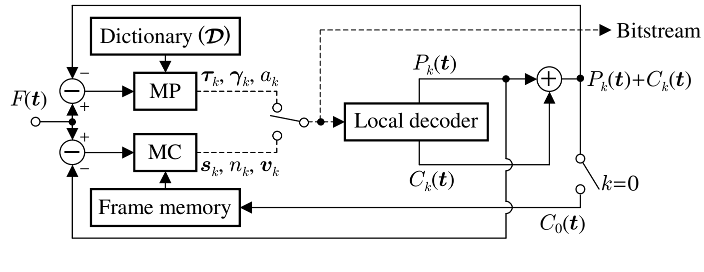

Figure 2: Block diagram of the encoder.

4.2 Determination of MP Parameters

The MP algorithm described in Section 2 is applied to the residual signal $e_{k-1}(\boldsymbol{t}) = F(\boldsymbol{t}) - \big(P_{k-1}(\boldsymbol{t}) + C_{k-1}(\boldsymbol{t})\big)$. In this paper, the following cost J_P is used for determination of the MP parameters of $\boldsymbol{\tau}_k$, $\boldsymbol{\gamma}_k$ and a_k:

$$J_P = \Delta SSE / \Delta R = a_k^2 \,/\, R_P(\boldsymbol{\tau}_k, \boldsymbol{\gamma}_k, a_k), \qquad (7)$$

where $\Delta R = R_P(\boldsymbol{\tau}_k, \boldsymbol{\gamma}_k, a_k)$ is a coding rate for encoding the parameter set $\{\boldsymbol{\tau}_k, \boldsymbol{\gamma}_k, a_k\}$. By maximizing the cost J_P instead of a_k^2 in Eq. (4), we can obtain a more appropriate set of these parameters in terms of rate-distortion and can surely improve coding efficiency when $R_P(\boldsymbol{\tau}_k, \boldsymbol{\gamma}_k, a_k)$ is not a constant for all of the combination of the parameters. In the coding scheme, the shift parameter $\boldsymbol{\tau}_k$ is represented with integer-pel accuracy and is encoded using a run-length coding method with a variable-length code (VLC). The index $\boldsymbol{\gamma}_k$ is simply expressed by a fixed-length code (FLC), and the coefficient a_k is quantized by a linear quantizer with a dead-zone and is encoded using another VLC. In addition, we employ the 2-D separable Gabor functions used in [3] as the dictionary $\boldsymbol{\mathcal{D}}$.

4.3 Determination of MC Parameters

The MC parameters $\boldsymbol{s}_k$, n_k and $\boldsymbol{v}_k$ are determined by the iterative MC method proposed in Section 3. Here, if the MC method is adopted in the current stage, the signal $P_k(\boldsymbol{t})$ reproduced by MP is unchanged. And in this case, a value of ΔSSE is expressed as:

$$\begin{aligned}\Delta SSE = {} & \big\|\big(F(\boldsymbol{t}) - P_{k-1}(\boldsymbol{t})\big) - C_{k-1}(\boldsymbol{t})\big\|^2 \\ & - \big\|\big(F(\boldsymbol{t}) - P_k(\boldsymbol{t})\big) - C_k(\boldsymbol{t})\big\|^2. \end{aligned} \qquad (8)$$

Comparing the above equation with Eq. (6), it turns out to be sufficient to execute the MC method by regarding the signal $\big(F(\boldsymbol{t}) - P_{k-1}(\boldsymbol{t})\big) = \big(F(\boldsymbol{t}) - P_k(\boldsymbol{t})\big)$ as the original one instead of $F(\boldsymbol{t})$. Consequently, an appropriate set of the parameters $\{\boldsymbol{s}_k, n_k, \boldsymbol{v}_k\}$ is also determined in the almost same way as the MP method by maximizing the following cost J_C for the MC method:

$$\begin{aligned} J_C &= \Delta SSE/\Delta R \\ &= \sum_{\boldsymbol{t} \in \boldsymbol{B}(\boldsymbol{s}_k, n_k)} \Big\{ \big(\big(F(\boldsymbol{t}) - P_{k-1}(\boldsymbol{t})\big) - C_{k-1}(\boldsymbol{t})\big)^2 \\ &\qquad - \big(\big(F(\boldsymbol{t}) - P_{k-1}(\boldsymbol{t})\big) - C_0(\boldsymbol{t} - \boldsymbol{v}_k)\big)^2 \Big\} \\ &\qquad\qquad / \, R_C(\boldsymbol{s}_k, n_k, \boldsymbol{v}_k), \end{aligned} \qquad (9)$$

where $\Delta R = R_C(\boldsymbol{s}_k, n_k, \boldsymbol{v}_k)$ is a coding rate for encoding the parameter set $\{\boldsymbol{s}_k, n_k, \boldsymbol{v}_k\}$. The motion vector $\boldsymbol{v}_k$ is represented with half-pel accuracy and is encoded using the same VLC as utilized in the H.263 [5], while the other two parameters are encoded using FLCs. As a matter of fact, maximization of the cost J_C is carried out only for two parameters $\boldsymbol{s}_k$ and n_k, because the motion vector $\boldsymbol{v}_k$ which maximizes ΔSSE in each block can be easily detected in advance as mentioned in Section 3. Though mixed use of two kinds of costs J_C and ΔSSE seems to be somewhat inconsistent, this simplifies the coding operation and reduces computational load remarkably. In this case, search for the motion vector $\boldsymbol{v}_k$ is needed again only in the blocks where the signal $P_k(\boldsymbol{t})$ is renewed by the MP method.

4.4 Adaptive Selection of MP and MC

By comparing values of the two costs J_P and J_C, a more appropriate operation in a rate-distortion sense is alternatively selected from between MP and MC in each stage as follows. These two values indicate the respective maximum ones of J_P and J_C.

- If $J_P > J_C$,

$$P_k(\boldsymbol{t}) = P_{k-1}(\boldsymbol{t}) + a_k \cdot g_{\boldsymbol{\gamma}_k}(\boldsymbol{t} - \boldsymbol{\tau}_k) \qquad (10)$$

$$C_k(\boldsymbol{t}) = C_{k-1}(\boldsymbol{t}), \qquad (11)$$

- else,

$$P_k(\boldsymbol{t}) = P_{k-1}(\boldsymbol{t}) \qquad (12)$$

$$C_k(\boldsymbol{t}) = \begin{cases} C_0(\boldsymbol{t} - \boldsymbol{v}_k) & (\boldsymbol{t} \in \boldsymbol{B}(\boldsymbol{s}_k, n_k)) \\ C_{k-1}(\boldsymbol{t}) & (\text{otherwise}). \end{cases} \qquad (13)$$

Then a value (k) of the counter increases by one, and the iterative coding operation is repeated until a coding rate R in each frame reaches a target rate R_T as shown in Figure 3. This simple method for rate control can accurately realize CBR coding with very small buffer size. On the other hand, VBR coding can be also realized by introducing the following condition to finish the iterative coding operation:

$$\text{if} \qquad \max\{J_P, J_C\} < \lambda, \qquad \text{finish}, \qquad (14)$$

where λ is a constant which corresponds to the Lagrangian multiplier for optimization in a rate-distortion sense. In this rate control method for VBR coding, the coding operation for each frame finishes just after $\Delta\text{SSE}/\Delta R$ becomes less than λ as illustrated in Figure 3.

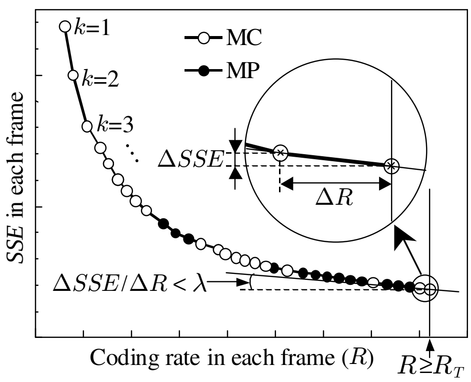

Figure 3: Rate control methods for CBR and VBR coding.

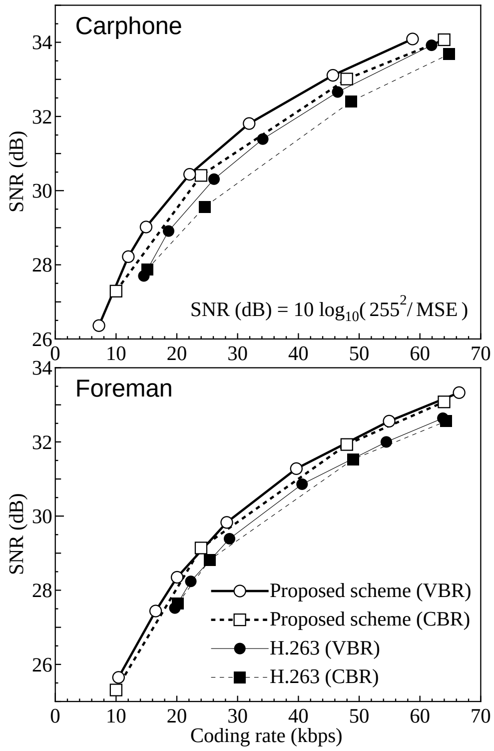

Figure 4: Coding performance.

5 SIMULATION RESULTS

In order to evaluate the performance of the proposed coding scheme, computer simulation is carried out. Luminance signals of test sequences called ‘Carphone’ and ‘Foreman’ (QCIF, 10 frames/second, 10 seconds) are used for the simulation. Figure 4 shows Rate-SNR curves of the proposed scheme and the H.263 standard [5]. In the proposed scheme, the JPEG baseline system is used for the first frame as an intra coding method. On the other hand, the results of the H.263 are obtained through UBC’s TMN 3.2 encoder software [6]. VBR coding of the H.263 is simulated with fixing a value of the quantization parameter (QP) for each frame. The figure demonstrates that the proposed scheme outperforms the H.263 by 0.3–0.9 dB in SNR of CBR coding and by 0.4–1.0 dB in SNR of VBR coding respectively. In the case of CBR coding, the proposed scheme exactly controls a coding rate R in each frame so that $|R - R_T| \,/\, R$ can be always less than 8 %. On the contrary, it often exceeds 60 % in the H.263. Furthermore, a coding rate of the MC parameters in the proposed scheme generally increases in proportion to the target rate R_T, while that of motion vectors in the H.263 is almost constant regardless of a value of R_T. These facts indicate that the proposed scheme can control a coding rate freely and can realize appropriate bit-allocation to both MC and MP according to the target rate. Figure 5 shows examples of reconstructed images. Though the proposed scheme employs a kind of block-based MC, the blocking artifacts which are

(a) Proposed scheme (29.4 dB) (b) H.263 (27.0 dB)

Figure 5: Examples of reconstructed images of ‘Carphone’ at 24 kbps of CBR coding (frame no. = 250).

noticeable in the reconstructed image of the H.263 diminish certainly in that of the proposed scheme. This is not only because the MP algorithm is free from such artifacts in nature, but also because the proposed scheme determines positions and size of blocks flexibly and appropriately so that their boundaries can fit contours of moving objects.

6 CONCLUSIONS

This paper has proposed a video coding scheme based on seamless combination of MC and MP. In this scheme, the iterative operations of MC and MP are carried out alternatively and a more suitable one is adaptively selected in terms of rate-distortion. Simulation results demonstrate that such collaboration of MC and MP can enjoy the merits of both operations and provides better coding performance than the conventional hybrid coding scheme in both CBR and VBR coding.

Another feature of the proposed coding scheme is its simple structure in the decoding process. A decoder which is experimentally implemented by using the Java language can run on most of popular Web browsers in real-time. Moreover, utilizing the JPEG decoder included in the Java generic library for intra-frames, entire file size of the compiled program for the decoder is reduced to about 12 kbytes. These facts are beneficial to cross-platform multimedia applications where a decoder itself is downloaded through networks on demand.

References

[1] Y. W. Lee et al., “Towards MPEG4: An Improved H.263-Based Video Coder” Signal Processing.: Image Comm., Vol.10, Nos.1–3, pp.143–158, July 1997.

[2] R. Neff et al., “Very Low Bit-Rate Video Coding Based on Matching Pursuits”, IEEE Trans. on CSVT, Vol.7, No.1, pp.158–171, Feb. 1997.

[3] M. R. Banham et al., “A Selective Update Approach to Matching Pursuits Video Coding”, IEEE Trans. on CSVT, Vol.7, No.1, pp.119–129, Feb. 1997.

[4] S. Mallat et al., “Matching Pursuits with Time-Frequency Dictionaries”, IEEE Trans. on SP, Vol.41, No.12, pp.3397–3415, Dec. 1993.

[5] ITU-T SG15, “Video Coding for Low Bitrate Communication”, Draft ITU-T Rec. H.263, May 1996.

[6] Signal Processing and Multimedia Group, University of British Columbia, “TMN (H.263+) encoder version 3.2”.