A STUDY ON PRIORITY-BASED SCALABLE CODING

Ryoichi Kawada, Atsushi Koike, Masahiro Wada and Shuichi Matsumoto

Visual Communications Laboratory
KDD R&D Laboratories
2-1-15 Ohara, Kamifukuoka, Saitama 356-8502, Japan
E-mail: kawada@kddlabs.co.jp

ABSTRACT

Recently, the concept of object scalability that also takes the contents of pictures into account unlike conventional scalability, has been proposed in MPEG-4. Object scalability in MPEG-4, however, assumes a VOP which requires the definition of shapes, so that it is not applicable to conventional coding schemes such as MPEG-2. This paper proposes a method of object scalable coding that can be adopted in these conventional block-based coding schemes. This method is based on SNR scalability, and it realizes scalability by differently controlling quantization on prioritized and non-prioritized parts. The bit-stream is compatible with SNR scalability, and for the enhancement layer picture, a similar SN ratio to that in normal SNR scalability can be obtained.

1 INTRODUCTION

Conventionally, as an example of scalable coding schemes in picture coding, spatial / SNR / temporal scalability has been adopted in MPEG-2, etc. On the other hand, the concept of object scalability has been proposed in MPEG-4 recently. This, as opposed to conventional scalability, intends to realize a scalability that considers even the contents of pictures, and it defines layers based on the priority of each object. Its application includes picture query and the expectation is large.

For example, there have been studies on conspicuousness that investigate the part in the picture that viewers tend to look at[1][2]. The priority might be placed upon such conspicuous parts in the picture to perform scalable coding.

The present object scalability, however, assumes the VOP of MPEG-4, which requires the definition of shapes. Therefore, it cannot be realized in schemes such as the conventional MPEG-2 that do not have such concepts.

The objective of this paper is therefore to obtain a method of priority-based object scalable coding that can be adopted in these conventional block-based coding schemes.

In the rest of the paper, Section 2 firstly describes the utilization of SNR scalability for the realization of this objective. Next, it shows that this method can provide a similar SN ratio for the enhancement layer picture to that of normal SNR scalability, by using the Lagrange multiplier method and dynamic programming. It also mentions a means of extension to multi-layer processing, as well as some examples of how to determine the priority. Section 3 shows the effectiveness of our method for object scalability by actually applying it to picture coding in computer simulation experiments. Section 4 concludes the paper.

2 PRIORITY-BASED SCALABLE CODING SCHEME

2.1 Use of SNR Scalability

We consider utilizing SNR scalability as a method for realizing a scalable coding that is based on the priority in the picture.

For simplicity, let us consider the case of two layers. The principle of SNR scalability is that the coding noise signal in the base layer picture (i.e., the difference from the original picture) is coded and transmitted in the enhancement layer. Thus, under the scheme of SNR scalability, we can realize a scalability that is based on objects as follows.

- In the base layer, fine quantization is applied for prioritized parts in the picture, and coarse quantization is applied for the other parts.

As a result, the receiver side can, by receiving the base layer, decode only the prioritized parts with high picture quality, and obtain the other parts with lower SN ratio. By decoding the enhancement layer also, it obtains the whole picture with high quality.

2.2 Constraints on Base Layer Coding

It is desirable that, even after the above processing, the final quality of the enhancement layer pictures should not be inferior to that in the case of normal SNR scalable coding. The following examines what constraints on the base layer coding are required.

Let us consider that the picture to be coded consists of N macroblocks(MB's). Let s_i denote the signal variance of the i-th MB ($i = 1, \cdots, N$). When the coding noise power of each MB is σ_i^2 after base layer coding, this σ_i^2 is considered the variance of the signal to be coded in the enhancement layer. Let n_i denote the power of the coding noise in each MB after enhancement layer coding.

Now, the objective is the minimization of noise in the enhancement layer picture, i.e., $\sum_{i=1}^{N} n_i \rightarrow$ min.

The constraint is that each generated entropy for the base layer and the enhancement layer is fixed, i.e.,

$$\sum_{i=1}^{N} \log_2 \frac{\sigma_i^2}{n_i} = R_E, \quad \sum_{i=1}^{N} \log_2 \frac{s_i}{\sigma_i^2} = R_L \qquad (1)$$

where $s_i \geq \sigma_i^2, \ \ \sigma_i^2 \geq n_i \ \ (\forall i)$.

This is solved by using the Lagrange multiplier method. The Lagrangian is as follows.

$$\begin{aligned} & L(\{\sigma_i^2\}, \{n_i\}, \lambda_L, \lambda_E, \{\mu_{Li}\}, \{\mu_{Ei}\}) \\ & = \lambda_L(R_L - \textstyle\sum_i \log_2 \frac{s_i}{\sigma_i^2}) + \lambda_E(R_E - \sum_i \log_2 \frac{\sigma_i^2}{n_i}) \\ & + \textstyle\sum_i \mu_{Li}(s_i - \sigma_i^2) + \sum_i \mu_{Ei}(\sigma_i^2 - n_i) - \sum_i n_i \end{aligned} \qquad (2)$$

By solving this we obtain the following (the derivation is omitted).

$$\mu_{Ei} = \mu_{Li} \equiv \mu_i \ \ (\forall i), \quad \lambda_E = \lambda_L \equiv \lambda \qquad (3)$$

$$n_i = \lambda / \ln 2 \ \text{ or } \ n_i = \sigma_i^2 = s_i \ \ (\forall i) \qquad (4)$$

In other words, for every MB, the same degradation is introduced or total discarding is performed. Furthermore, It can be said that the MB that is totally discarded in the enhancement layer is also discarded in the base layer.

As shown above, for the enhancement layer, specific guidelines (quantization with the same step size for every MB) are given, but none for the base layer. They are necessary conditions, but it is expected that the base layer coding will have a rather large degree of freedom.

2.3 Preliminary Experiments

Thus, in order to confirm the above, we try this optimization by using dynamic programming for a set of s_i obtained from an actual picture sequence.

Stage t: the t-th macroblock MB_t

State x_t: the pair (x_{Lt}, x_{Et}), where $x_{Lt} = \sum r_{Li}$ is the sum of $r_{Lt}, r_{L(t+1)}, ..., r_{LN}$, the base layer entropy that is consumed by macroblocks $MB_t, MB_{t+1}, ..., MB_N$, and $x_{Et} = \sum r_{Ei}$ is the sum of $r_{Et}, r_{E(t+1)}, ..., r_{EN}$, the enhancement layer entropy that is consumed by them.

Decision k: a pair (k_L, k_E), where k_L is a candidate for base layer entropy that may be consumed by MB_t, and k_E is that of enhancement layer entropy.

Cost function $f_t(x_t)$: the minimum value of $\sum n_i$, the total power of coding noise in the enhancement layer for $MB_t, MB_{t+1}, ..., MB_N$ at the state x_t.

The optimality equation is Eq.(5) below.

$$\begin{cases} f_N(x_N) = s_N 2^{-(r_{LN}+r_{EN})} \\ f_t(x_t) = \min_{\forall k_L \leq x_{Lt}, \forall k_E \leq x_{Et}} \\ \qquad \{s_t 2^{-(k_L+k_E)} + f_{t+1}(x_t - k)\} \\ \qquad\qquad (2 \leq t \leq N-1) \\ f_1(R_L, R_E) = \min_{\forall k_L \leq R_L, \forall k_E \leq R_E} \\ \qquad \{s_1 2^{-(k_L+k_E)} + f_2((R_L, R_E) - k)\} \end{cases} \qquad (5)$$

Also, for comparison, an optimization that is based on each layer separately is performed. In other words, by using dynamic programming, firstly the base layer coding is optimized, and after that, the enhancement layer coding is optimized, the object of which is the coding noise in the base layer. In this case, the optimality equation of the base layer coding is Eq.(6) below.

Stage t: the t-th macroblock MB_t

State x_t: $\sum r_{Li}$, the sum of r_{Lt}, $r_{L(t+1)}$,..., r_{LN}, which is the base layer entropy that is consumed by macroblocks MB_t, MB_{t+1},..., MB_N.

Decision k: k_L, a candidate for base layer entropy that may be consumed by MB_t.

Cost function $f_t(x_t)$: the minimum value of $\sum \sigma_i^2$, the total power of coding noise in the base layer for MB_t, MB_{t+1},..., MB_N at the state x_t.

$$\begin{cases} f_N(x_N) = s_N 2^{-(r_{LN})} \\ f_t(x_t) = \min_{\forall k_L \leq x_{Lt}} \{s_t 2^{-k_L} + f_{t+1}(x_t - k)\} \\ \qquad\qquad (2 \leq t \leq N-1) \\ f_1(R_L) = \min_{\forall k_L \leq R_L} \{s_1 2^{-k_L} + f_2(R_L - k)\} \end{cases} \qquad (6)$$

Next, using the resultant coding noise σ_i^2 of the base layer, the enhancement layer coding is optimized. The optimality equation is Eq.(7) below.

Stage t: the t-th macroblock MB_t

State x_t: $\sum r_{Ei}$, the sum of r_{Et}, $r_{E(t+1)}$,..., r_{EN}, which is the enhancement layer entropy that is consumed by macroblocks MB_t, MB_{t+1},..., MB_N.

Decision k: k_E, a candidate for enhancement layer entropy that may be consumed by MB_t.

Cost function $f_t(x_t)$: the minimum value of $\sum n_i$, the total power of coding noise in the enhancement layer for MB_t, MB_{t+1},..., MB_N at the state x_t.

Tbl. 1: Results of optimization by DP. Left: optimized over both layers. Right: each layer optimized separately.

MB#	s_i	σ_i^2	n_i	σ_i^2	n_i
00	426.9	43.0	15.2	55.1	15.2
01	106.9	18.5	15.2	54.6	15.2
02	119.8	54.9	15.3	54.9	15.3
03	376.2	47.0	15.2	54.7	15.2
04	269.5	73.5	15.3	54.8	15.3
05	183.7	78.9	15.2	54.6	15.2
06	216.8	116.9	15.3	54.8	15.3
07	184.8	100.8	15.3	54.9	15.3
08	81.9	37.1	15.3	54.8	15.3
09	159.6	87.0	15.2	54.6	15.2
10	113.7	52.2	15.3	55.1	15.3
11	21.0	17.1	15.2	20.8	15.2
12	19.5	16.4	15.3	19.3	15.3
13	21.8	18.9	15.2	21.5	15.2
14	114.7	62.5	15.3	54.9	15.3
15	118.8	96.7	15.3	55.0	15.3
16	352.5	62.3	15.2	54.7	15.2
17	211.4	109.2	15.2	54.6	15.2
18	177.5	98.9	15.2	55.1	15.2
19	124.1	78.7	15.3	55.1	15.3
20	96.3	89.3	15.3	54.8	15.3
21	154.6	131.4	15.2	54.7	15.2
22	71.4	55.1	15.3	55.1	15.3
23	71.5	62.8	15.2	54.6	15.2
24	45.8	43.4	15.2	45.3	15.2
25	60.2	56.4	15.2	54.6	15.2
26	109.7	93.3	15.3	54.9	15.3
27	380.2	45.5	15.2	54.7	15.2
28	452.7	52.5	15.3	54.8	15.3
29	184.8	46.7	15.3	54.9	15.3
30	341.1	51.3	15.2	54.7	15.2
31	608.9	66.1	15.3	55.0	15.3
32	724.3	79.5	15.3	55.0	15.3
33	692.4	76.0	15.3	54.9	15.3
34	680.7	75.5	15.2	54.6	15.2
35	94.4	15.5	15.3	54.9	15.3
36	7.3	7.2	7.2	7.2	7.2
37	7.4	7.3	7.2	7.3	7.2
38	14.5	14.3	14.1	14.3	14.1
39	726.6	40.3	15.2	54.6	15.2
40	1338.3	142.2	15.3	54.8	15.3
41	224.0	25.7	15.3	54.8	15.3
42	40.8	15.4	15.2	40.3	15.2
43	299.8	23.5	15.3	54.7	15.3
44	363.1	36.5	15.2	54.6	15.2
45	14.1	13.9	13.8	13.9	13.8
46	11.5	11.4	11.3	11.4	11.3
47	87.5	15.5	15.3	54.9	15.3
48	195.3	18.0	15.3	55.0	15.3
49	182.1	105.9	15.2	54.7	15.2
Ave.	233.6	55.8	14.8	47.9	14.8

$$\begin{cases} f_N(x_N) = \sigma_N^2 2^{-r_{EN}} \\ f_t(x_t) = \min_{\forall k_E \le x_{Et}} \{\sigma_t^2 2^{-k_E} + f_{t+1}(x_t - k)\} \\ \qquad\qquad\qquad\qquad (2 \le t \le N-1) \\ f_1(R_E) = \min_{\forall k_E \le R_E} \{\sigma_1^2 2^{-k_E} + f_2(R_E - k)\} \end{cases} \tag{7}$$

The number of MB's is 50. As a computation result, Tbl.1 shows the coding noise for each MB and the average over all MB's.

In case of separate optimization for each layer, the same degradation is introduced or total discarding is performed (when the power of the original signal is smaller than the degradation) for each MB (the calculation error is about ± 0.1). Also, between the case of integrated optimization over the two layers and the case of separate optimization for each layer, the coding noise in the enhancement layer is identical.

In other words, the minimum $\sum n_i$ can be obtained by optimizing the coding of each layer separately (quantization by the same step size).

It can also be said that, based on the resultant σ_i^2, we can change them as far as they satisfy $\prod \sigma_i^2$ = constant that is derived from the condition Eq.(1), and $s_i \ge \sigma_i^2 \ge n_i$.

For example, when N is set at 2 and separate optimization for each layer has obtained $\sigma_1^2 = 10$ and $\sigma_2^2 = 10$, adjusting the priority of quantization to get $\sigma_1^2 = 1$ and $\sigma_2^2 = 100$ does not lead to degradation of quality of the enhancement layer picture as long as the above-mentioned conditions are satisfied.

Therefore, if this adjustment is performed according to the parts to be prioritized, the object scalability based on the priority in the picture can be realized, which does not need overheads as compared with normal SNR scalability.

2.4 Extension to Multi-layered Processing

This priority-based scalable coding, like SNR scalable coding, can be extended to multi-layered processing. This can be described as follows.

In the first layer (base layer), the most prioritized parts are decoded with high quality. If the second layer bit-stream is also received, in addition to the most prioritized parts, the second-most prioritized parts are decoded with high quality. If all the layers are received, the whole picture is decoded with high quality.

In order to define the degree of priority inside each layer, we set a priority function, which influences the quantization control for each MB, on the picture plane.

Let the priority function of the k-th layer, $f_k(x, y)$, represent the parts that are prioritized in the first through the k-th layers. If we have n layers, the priority function for the highest layer is supposed to have a constant (maximal) value over the whole picture plane. Let the value be 100 for example. Then the following can be assumed.

$$f_n(x, y) = 100 \quad \forall(x, y)$$

$$f_k(x, y) \geq f_{k-1}(x, y) \quad \forall(x, y) \quad (1 < k \leq n)$$

When quantization is performed in each layer, for the parts where the priority function takes larger values, finer quantization is performed.

2.5 Examples for Determination of Degree of Priority

For determination of degree of coding priority, different methods can be used. Thus in the computer simulation experiments of Section 3, the following methods are utilized.

• Determination according to the magnitude of intra-frame variance in each MB.

• Determination according to the magnitude of conspicuousness on the picture plane.

• Prioritization of automatically extracted foreground.

Each method is described below.

2.5.1 Determination according to the magnitude of intra-frame variance in each MB

If we want to prioritize the parts that a viewer tends to look at, this method can be used for simplicity.

It is based on a simple assumption that, in the picture, the parts that contain complex texture rather than flat parts have more information, so that viewers will tend to look at them.

For each macroblock, the intra-frame variance there is calculated, and the larger value it takes, the finer quantization is performed.

2.5.2 Determination according to the magnitude of conspicuousness on the picture plane

More detailed research has been conducted about how viewers tend to look at the parts in the picture. For example, reference [1] defines the conspicuousness in the picture according to the following algorithm.

(1) Performs segmentation of the picture.

(2) For each segment, calculates a set of features that include color hue, color saturation, brightness, color difference from neighboring segments, and the area of the segment.

(3) Determines the magnitude of conspicuousness of the segment from the features by referring to a table.

In this study, the priority is defined for each MB, so that by modifying the segment boundaries to be along the MB's boundaries, we can use the above method to determine the degree of priority.

2.5.3 Prioritization of automatically extracted foreground

For purposes such as visual surveillance, schemes for foreground extraction by background difference methods have been studied. For example, in reference [3], the following method is utilized.

(1) Background modeling
Performs pixel-by-pixel processing. Brightness, saturation and hue are used, and it is assumed that the values in the background follow a Gaussian distribution.

(2) Determination of foreground pixels
Judges that a pixel belongs to the foreground in the present frame when its value is not within twice the standard deviation from the average of the current background value. (Saturation is processed exceptionally.)

(3) Update of the background model
Updates the current background model by simple exponential smoothing for each background pixel in the present frame.

In this study, the priority is set per MB, so that by grouping the above foreground pixels so as to make the boundaries to be along macroblocks, we can define the foreground as prioritized parts.

2.6 Features of the Proposed Scheme

As shown above, the features of this object scalable coding scheme are as follows.

• The bit-stream is compatible with that of SNR scalable coding.

• By adequately defining prioritized parts on the picture plane, a SN ratio that is similar to that of normal SNR scalability can be obtained in the enhancement layer.

• It can be extended to a multi-layered scheme with more than two layers like SNR scalable coding.

• For defining prioritized parts, different algorithms can be utilized. (Defining them manually is also possible.)

3 COMPUTER SIMULATION EXPERIMENTS

In the proposed scheme, the ideal processing should start with the optimal coding in each layer as described before. However, this processing would be very complex, so that here for simplicity, the following method is applied to coding experiments using real pictures.

First, by the conspicuous method of 2.5.2, the prioritized area is defined, which is set at about 40% of the whole. Then for the non-prioritized area in the base layer, MQUANT, the quantization parameter in MPEG-2, is set at the maximal value so as to truncate the coefficients. In the prioritized area, the normal TM5 rate control is applied. In the enhancement layer, the whole picture is encoded under normal rate control. Step 3 in the TM5 rate control is omitted.

For comparison, the normal SNR scalable coding is also applied, where in the base layer encoding is performed by TM5 and the resultant coding noise is encoded in the enhancement layer.

Six titles of test picture sequences, cheerleaders, flamingoes, green leaves, marching in, mobile & calendar and soccer are used, and the coding rate is 5.6Mbps for each layer.

Tbl.2 shows PSNR's of the prioritized area in the base layer decoded picture and those of the whole area in the enhancement layer decoded picture.

Tbl. 2: PSNR's of decoded pictures [dB]. Pri: priority-based scalability. Snr: normal SNR scalability.

Pictures	Base, prioritized area			Enhance, whole	
	Pri	Snr	increase	Pri	Snr
Cheer.	34.9	32.2	+2.7	35.0	34.8
Flamin.	35.0	33.5	+1.5	34.7	34.5
Green.	32.0	28.6	+3.4	31.5	30.7
March.	31.3	29.3	+2.0	31.0	30.7
Mobile	34.9	30.2	+4.7	32.2	31.5
Soccer	32.4	29.5	+2.9	31.0	30.4

From this result we can confirm that, in the proposed scalable coding which prioritizes conspicuous parts, the prioritized area in the base layer can be coded so as to give higher PSNR's than those in the normal SNR scalable coding by about 1.5-4.7 dB. At the same time, PSNR's in the enhancement layer are also about the same or rather higher than those of SNR scalability. It can be said that the encoding of normal SNR scalability, although each layer is separately encoded, is not necessarily optimal.

Next, in order to investigate the difference due to the methods for defining the prioritized area, an experiment on the proposed scalable coding was conducted by using each of the aforementioned three methods for defining prioritized area, with the picture "race circuit". With all the methods, parameters were adjusted so as to set the prioritized area to about 40% of the whole picture. Tbl.3 shows the PSNR's of the prioritized area in the base layer decoded picture and the PSNR's of the whole enhancement layer decoded picture. The coding rate is 5.6 Mbps for each layer. From this result we can confirm that with each of the methods for definition of prioritized parts, the proposed scalable coding provides, higher PSNR's than the normal SNR scalable coding in the prioritized area in the base layer by 1.0-2.9 dB. Also, PSNR's of the enhancement layer decoded pictures, as in Tbl.2, are rather higher with the proposed scalable coding than with the normal SNR scalable coding.

Tbl. 3: PSNR's of decoded pictures [dB]. Pri: priority-based scalability. Method 1: priority by variance. Method 2: priority by conspicuousness. Method 3: priority by foreground extraction. Snr: normal SNR scalability.

Method	Base, prioritized area			Enhance, whole	
	Pri	Snr	increase	Pri	Snr
1	39.3	36.4	+2.9	40.9	40.4
2	38.6	37.6	+1.0	40.7	40.4
3	39.0	37.6	+1.4	40.6	40.4

Incidentally, the reason why, in the case with the priority definition method 1 (the method based on intra-frame variance), the PSNR's of the prioritized area of the base layer picture in the normal SNR scalable coding scheme are lower than those in the cases with the other definition methods, is as follows. In general, the area with larger values of intra-frame variance does not contain flat parts and has a lot of information to be encoded (Fig.1). On the other hand, the areas defined by the priority definition methods 2 and 3 contain many flat parts, and less information has to be encoded (Fig.2,Fig.3) than in the area defined by the first method.

The above discussion has shown that even with

Fig. 1: The prioritized area defined by using intra-frame variance. The black parts are non-prioritized.

Fig. 2: The prioritized area defined by using conspicuousness. The black parts are non-prioritized.

the simple method utilized in this experiment, a priority-based object scalable coding can be effectively realized.

In the future, even better performance can be expected by improving the method of defining the prioritized parts.

4 CONCLUSIONS

In this paper, in order to realize object scalable coding that is applicable to conventional block-based coding such as MPEG-2, the authors proposed a scheme where scalability is realized by quantization control according to the priority on the picture surface, which is based on SNR scalability.

It realizes the object scalability by performing finer quantization in more prioritized parts in the base layer of SNR scalable coding. For defining the priority, different methods such as the conspicuousness method can be used.

The bit stream of this scheme is compatible

Fig. 3: The prioritized area defined by extracting the foreground. The black parts are non-prioritized.

with that of SNR scalability. Also, the enhancement layer picture can yield a SN ratio similar to that of normal SNR scalable coding. An extension to multi-layer processing is also easily made.

The authors hope that this study will help to enlarge the application area of scalable coding schemes, such as video query.

Acknowledgment

The authors would like to thank Dr. Akiba and Dr. Asami of KDD R&D Laboratories for their valuable suggestions and encouragement. This research has been conducted within the context of "Research and development on ISDB technologies", contract research from the Telecommunications Advancement Organization of Japan.

References

[1] S.Tanaka, et al., "A method for extracting attractive regions based on physical features contributing to attractiveness," Journal of ITE, 52, 6, pp.881-890, Jun.1998.

[2] L.Itti, et al., "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. PAMI-20, 11, pp.1254-1259, Nov.1998.

[3] A.R.J. A.R.J.Francois, et al.,"Adaptive color background modeling for real-time segmentation of video streams," Proc. Inter. Conf. Image Science,Systems,and Technology (1999)