

A HYBRID APPROACH WITH DRIFT MINIMIZATION AND CORRECTION FOR A LAYERED H.263 CODEC

Sheng-Jin Hsieh, Ja-Ling Wu, Senior Member, IEEE,
Chien-Wu Tsai, and Jyh-Liang Dai
Communications and Multimedia Lab.
Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan, R.O.C.

Abstract—Layered video coding allows the video information to be transmitted in multiple video bitstreams to achieve scalability. And the quality of the base layer is limited by drift. In this paper, we propose a hybrid approach to eliminate the drift in an H.263-based video codec with two-layer scalability. The base layer decoder takes advantage of the minimal drift approach to obtain the reconstructed services that have the minimum drift. Then, the encoder sends an additional drift correction signal, which is used to eliminate drift completely, to the decoder. In this way, the proposed approach can eliminate the drift completely and reduce the overhead required for the drift correction.

Index Terms—Drift correction, layered video coding, H.263 coding.

I. Introduction

The layered video coding can provide the scalabilities of spatial, temporal, and quality in many applications, e.g., multipoint videoconference, video services in ATM network, program deliveries of traditional TV and HDTV, etc. Within the syntax of MPEG-2 video coding standard [1], there are four layered coding techniques. A major drawback of layered coding is that the overall coding overhead will increase for high quality service. And this effect would violate the low bitrate requirement in the application of videoconference.

In the past researches, the layered coding using bitstream decomposition has been shown to be efficient from the viewpoints of implementation complexity and coding overhead. However, since the base layer does not contain information about high-frequency components, the service quality of the reconstructed video for the base layer decoder will heavily degrade over time until a refresh is activated. This effect is known as drift [2].

In this paper, we propose a hybrid approach to eliminate the drift in an H.263-based video codec with two-layer scalability. The base layer decoder takes advantage of the minimal drift approach [2,3] to obtain reconstructed services that have the minimum drift. Then, the encoder sends an additional drift correction signal to the decoder,

which is used to eliminate the drift completely. In this way, the proposed approach can eliminate the drift completely and reduce the overhead of the drift correction signal.

This paper is organized as follows. Section II briefly reviews the DCT-domain layered video coding techniques. Section III addresses the reasons why the drift is occurred and the methods how it can be eliminated. The implementation issues of the codec with drift-correction are investigated in Section IV. The experimental results are presented in Section V. Finally, conclusion and future work directions are given in Section VI.

II. Bitstream Decomposition

The layered coding divides the output of a single layer encoder into two or more frequency bands called the base layer and enhancement layers, respectively. The simplest structure of the layered coding is the two-layer structure, as shown in Fig. 1. A degraded service quality will be obtained by decoding only the information in the base layer. By adding the information contained in the enhancement layers, better service quality can be achieved.

For MPEG-2 coding standard [1], layered coding is referred as scalable coding. There are four kinds of scalability and each is applicable to various applications. The four kinds of scalability are as follows:

- (i) Data partitioning: It is also called bitstream decomposition. We conduct this method in our implementation. In transform or subband based coders, one can include the low-frequency coefficients or subsignals of low-frequency band in the base layer while leaving the high-frequency signal in the enhancement layers. In this way, the coding overhead is minimum as compared with other techniques [4].
- (ii) SNR scalability: Under the same spatial and temporal resolution, the base layer encodes the input signal with coarser quantizer, leaving the fine details quantized by finer quantizers specified in the enhancement layers.
- (iii) Spatial scalability: The base layer codes the input signal with subsampling, and the enhancement layers contain additional information for obtaining higher spatial resolution at the decoder.
- (iv) Temporal scalability: The base layer contains a bit stream with a lower frame rate, and the enhancement layers contain incremental information for obtaining an output with higher frame rates.

It is also possible to use combinations of aforesaid scalability schemes, which is called the hybrid scalability. Different layering methods, specified above, are suitable for different applications. There are three factors have to be considered before one can make a right choice among them. First, the increment of total data volume is the most important factor. No matter what kinds of layered coding schemes being used, there must have an increase in the total data volume for providing the higher quality service. Therefore, if the application needs the bitstreams from all layers, the total data volume may exceed the allowable budget. The second factor is the service quality provided by the base layer. The provided service quality of the base layer depends heavily on the distribution of the coding information among different layers in the adopted layered coding scheme. Finally, the complexity of implementation is the third factor that must be well considered. Using layered coding increases both the complexities of encoder and decoder. The increased complexity is resulted from two newly added components in the codec, i.e., a multiplexer for the decoder and a demultiplexer for the encoder. Since there is a trade-off between complexity and total data volume, it is impossible for an implementation to meet the optimal conditions for all the factors. That is, for a particular application, one can adjust the weighting of the above factors to make a best choice of layered coding schemes.

In the existing coding standards of videoconference, e.g. H.261 and H.263, the low bitrate demand is the most important thing to make them prevailing. If we want to apply the layered coding to videoconference applications, the adopted layered coding scheme must not increase the total bitrate too much. In this paper, we adopt the bitstream decomposition scheme for performing layered coding, not only for its minimal bitrate overhead but also for its less implementation complexity. An additional consideration besides bitrate and complexity is that the decoder should have the ability to eliminate the drift effect when drift correction information is added on.

The proposed encoding process is similar to that of the H.263, which uses the basic video coding techniques such as motion-compensation and DCT transformation. In interframe mode, the motion-compensated difference image is transformed using the DCT and then quantized. The quantized DCT coefficients are then scanned in a zigzag order, and finally the scanned coefficients are coded by using a combination of run-length and variable length coding.

The bitstream decomposition in the proposed two-layer coding scheme is performed during the zigzag scanning stage. We divide the DCT coefficients into two parts and called them the most significant part (MSP) and the least significant part (LSP), respectively. The MSP forms the base layer, while the LSP forms the enhancement layer. We propose two scanning patterns for the decomposition of DCT coefficients, denoted by P1 and P2, respectively (c.f. Fig. 2). The DCT coefficients in the top left 4x4 block of an

8x8 DCT block are coded as the base layer, and the remaining DCT coefficients are coded as the enhancement layer for scanning pattern P1. For scanning pattern P2, the first n DCT coefficients in the zigzag scanning order of an 8x8 block are coded as the base layer, and the remaining are coded as the enhancement layer. No matter what scanning pattern is chosen, the base layer contains almost the lowest frequency coefficients located at the top left corner of an 8x8 block. This DCT-domain decomposition possesses similar effect of performing subsampling on the DCT coefficients, it was called DCT subsampling in [3].

III. Drift Correction

A. Drift Effect

In the prescribed bitstream decomposition structure, motion compensation is performed on the original high-resolution video sequence, which includes both the low-frequency and high-frequency components generated by the encoder. But for the base layer decoding, motion compensation is performed on the low-resolution video sequence that contains the low-frequency components only. The discrepancy between the encountered information in the encoder and the base layer decoder implies that the motion compensation process cannot be identically replicated at the base layer decoder. This, of course, will cause an accumulation of errors in the decoded video sequence. After several frames are coded in interframe mode, the accumulated error can be sufficiently large to cause significant disparities between the decoded base layer service and a decimated version of the corresponding decoded full-resolution service. These resultant disparities are known as drift [2]. An example of the effect of drift is shown in Fig. 3.

There are two causes of drift. One is due to the use of suboptimal motion vectors, and the other is due to the DCT subsampling [3].

(i) Suboptimal motion vector

In the P1 bitstream decomposition scheme, the base layer decoder gets the decimated image (with halved size both in horizontal and vertical dimensions) after the lower 4x4 DCT coefficients are inverse transformed by a 4x4 IDCT. That is, the motion compensation is performed on quarter size of the original image at the base layer decoder. Thus, the motion vectors used in the base layer decoder are calculated by dividing the motion vectors obtained from the full-resolution encoder by two, both in horizontal and vertical directions. Some kinds of interpolation have been used to produce the subpixel accuracy so as to generate fractional motion vectors, the compensated result is not optimal and the errors still accumulate rapidly at the base layer decoder.

In the P2 bitstream decomposition scheme, the base layer decoder gets the lower 16 DCT coefficients in the zigzag scanning order and set the remaining higher 48 DCT coefficients to 0, then all these 64 DCT coefficients are inverse transformed by an 8x8 IDCT. In this way, the size

of the output image of the base layer decoder is the same as the original one. It is not necessary to do any adjustment of the motion vectors, and the drift effect of P1 will not occur in P2. Of course, the computational cost is a little higher in this case.

(ii) DCT subsampling

There are two reasons why DCT subsampling will introduce drift effect. First, the base layer decoder performs motion compensation only on low-frequency information from a previously decoded low-resolution image. But the encoder performs motion compensation on full-resolution decoded image which includes both the low-frequency and high-frequency information from a previously decoded full-resolution image. This disparity introduces drift in the decoded low-resolution service. Second, when displacement of motion vectors obtained in the full-resolution encoder is not a multiple of the block size, low-frequency information itself will introduce drift as well. This is because the operations of spatial shifting and DCT subsampling are not commutative. That is, the result of performing motion compensation and DCT subsampling in this order is not identical to the result obtained by performing in the reverse order.

From the mathematical derivation of drift given in [3], one can understand clearly about the relationships among drift, spatial shifting and DCT subsampling. Fig. 4 shows the relationship between an intraframe and a predicted frame for 1-D motion compensation. In this figure, b represents a predicted block in the predicted frame, h is the reference block of b in the intraframe obtained by motion estimation. f and g are the two blocks in the intraframe overlapping with h. That is, block h is across the block boundary between f and g, and they are divided into four parts f_1 , f_2 , g_1 , and g_2 , with f_2 and g_1 composing h.

To look upon the effect of DCT subsampling, h is transformed to the frequency domain. Let T be a real and orthogonal DCT matrix. Matrices H, F, and G denote the results after applying the transformation T on each block of h, f, and g, respectively. Assume that f, g, and T are all partitioned at the same position, and the upper part of each partition, has k rows, is used to indicate the motion vector there. It defines M(k) in eqn.(1) to represent the relationships among H, F, and G.

$$M(k) \equiv T_n \times n \begin{bmatrix} (T_2^t)_{(n-k) \times n} & 0 \\ 0 & (T_1^t)_{k \times n} \end{bmatrix}_{n \times 2n} \quad (1)$$

Fig. 5 shows the differences among them both in frequency and spatial domain. In the figure, the left half part shows the full-resolution encoder and the right half part shows the base layer decoder. Furthermore, the top half represents the operations on the spatial domain and the bottom half is on frequency domain. Symbol e in the figure represents the error block, and E, divided into E_1 and E_n , denotes the transformation result of e by transformation T. Symbol b represents the reconstructed block, and its transformation result B by T is consisted of B_1 and B_n . Symbol k represents

the 1-D motion vector obtained by motion estimation at the encoder side. Symbol k_s represents the 1-D motion vector used at the base layer side. The subscripts in T_n and T_s indicate the n-point and s-point DCT, respectively. Since DCT is a linear transformation, it has the outcome $B=H+E$. In practical implementation, the coefficients contained in F, G, and E have to be calculated. However, because the restoration of the error image always conducted in the spatial domain, there is no need to calculate the coefficients contained in H and B.

Due to DCT subsampling, the low-frequency coefficients are distributed to the base layer and high-frequency coefficients to the enhancement layer. F and G are divided into the low-frequency part and high-frequency part. The same denotation can be used in the base layer decoder, and it has $B_s=H_s+E$. Since an appropriate partition, such as $F_s=F_1$ and $G_s=G_1$, can be achieved by using DCT subsampling, it gets the result $H_s \neq H_1$, which demonstrates the existence of drift effect.

From the description of DCT subsampling, the low-frequency coefficients (obtained by performing n-point DCT) are inverse transformed by a s-point IDCT. Let $M^{(n)}(k)$ and $M^{(s)}(k_s)$ are defined in eqn.(1) with dimension $n \times 2n$ and $s \times 2s$, respectively. Then, the drift D due to DCT subsampling and spatial shifting can be expressed as

$$D \equiv [H_l - H_s] = \begin{bmatrix} M_{11}^{(n)}(k) - M_{11}^{(s)}(k_s) & M_{12}^{(n)}(k) \\ M_{13}^{(n)}(k) - M_{13}^{(s)}(k_s) & M_{14}^{(n)}(k) \end{bmatrix} \cdot \begin{bmatrix} F_l \\ F_h \\ G_l \\ G_h \end{bmatrix}, \quad (2.1)$$

where

$$H_l = \begin{bmatrix} M_{11}^{(n)}(k) & M_{12}^{(n)}(k) & M_{13}^{(n)}(k) & M_{14}^{(n)}(k) \end{bmatrix} \cdot \begin{bmatrix} F_l \\ F_h \\ G_l \\ G_h \end{bmatrix} \quad (2.2)$$

and

$$H_s = \begin{bmatrix} M_{11}^{(s)}(k_s) & M_{12}^{(s)}(k_s) \end{bmatrix} \cdot \begin{bmatrix} F_l \\ G_l \end{bmatrix}. \quad (2.3)$$

B. Drift Cancellation

There are many techniques have been proposed to limit drift effect in the decoded low-resolution image sequence [2][3]. Some of them do the job of drift minimization by the decoder, without modifying the encoder [2]. But doing this way, drift cannot be eliminated completely because the phenomenon of error accumulation caused by drift still exists. In order to achieve the goal of entirely eliminating the drift, it is necessary for the encoder to send an additional drift correction signal to the base layer decoder. The additional drift signal causes an important negative effect—the extra bitrate overhead, which one has to consider

seriously in the design. We will propose a hybrid method such that not only the drift can be eliminated completely, but also the bitrate overhead is kept as small as possible.

(i) Drift minimization

From eqn.(2), we know that the base layer decoder do not have any idea about the high-frequency information F_h and G_h . So, the error \hat{D} , defined in eqn.(3), cannot be cancelled at the base layer decoder.

$$\hat{D} = \begin{bmatrix} M_{11}^{(n)}(k) & M_{12}^{(n)}(k) & M_{13}^{(n)}(k) & M_{14}^{(n)}(k) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ F_h \\ 0 \\ G_h \end{bmatrix}. \quad (3)$$

To cancel the terms containing low-frequency parts in eqn.(2), we must set s equals to n so that the following conditions are met.

$$M_{11}^{(n)}(k) - M_1^{(s)}(k_s) = 0$$

and

$$M_{13}^{(n)}(k) - M_2^{(s)}(k_s) = 0$$

A simple method to satisfy these conditions is to let the base layer decoder performs motion compensation based on the full-resolution of the reconstructed image instead of the decimated one. To obtain the reconstructed image of full-resolution, we padded the remaining high-frequency coefficients by zeros in addition to the low-frequency coefficients contained in the base layer. After IDCT of the whole coefficients, we get the full-resolution reconstructed image in spatial domain. So, both the encoder and the base layer decoder must use the same n -point DCT in this method. With the full-resolution reconstructed image, we perform motion compensation on it and obtain the predicted image block as shown in eqn.(4).

$$\begin{bmatrix} M_{11}^{(n)}(k) & M_{12}^{(n)}(k) & M_{13}^{(n)}(k) & M_{14}^{(n)}(k) \\ M_{21}^{(n)}(k) & M_{22}^{(n)}(k) & M_{23}^{(n)}(k) & M_{24}^{(n)}(k) \end{bmatrix} \cdot \begin{bmatrix} F_l \\ 0 \\ G_l \\ 0 \end{bmatrix} \quad (4)$$

Again, taking only the low-frequency parts of eqn.(4) into account. Eqn.(5) represents the decoded image block after motion compensation.

$$\begin{bmatrix} M_{11}^{(n)}(k) & M_{12}^{(n)}(k) & M_{13}^{(n)}(k) & M_{14}^{(n)}(k) \end{bmatrix} \cdot \begin{bmatrix} F_l \\ 0 \\ G_l \\ 0 \end{bmatrix}. \quad (5)$$

Eqn.(3) shows which remaining part of the drift that did not eliminate by this method. In other words, this is one kind of drift minimization approaches performed only on the base layer decoder site, in which the encoder is unchanged.

(ii) Drift correction signal

The encoder calculates the drift as specified in eqn.(2)

and transmits this drift correction signal to the base layer decoder. When the base layer decoder receives the drift correction signal, the decoder can recover the drift error. In this way, the structure of the encoder has to be modified and the complexity of the encoder will be increased, too. The processing details of this approach will be described in Section IV.

(iii) Hybrid drift correction

This is a combined approach of the above two schemes. That is, the base layer decoder has the ability to minimize the overhead paid for drift correction, and the encoder has the ability to generate and transmit the required high-frequency information as the drift correction signal to the base layer decoder. From the previous derivation, we can retain the low-frequency information as specified in eqn.(5) for the base layer decoder. And the only drift correction signal needs to be transmitted is expressed in eqn.(3). Since this structure can effectively reduce the overhead that must be paid for sending drift correction signal, it is very suitable for the very low-bitrate videoconference applications. The details of the codec design will be presented in Section IV.

IV. The Design and Implementation Issues of the Proposed Codec

In this section, we will demonstrate three different two-layer codec structures. The codec structures are formed by the combinations of the bitstream decomposition techniques mentioned in section II and the drift correction methods specified in section III.

(A) Codec-1: P1 bitstream decomposition with drift correction

This codec structure was proposed by Mathew [2]. The encoder adopts the P1 scanning pattern to generate the layered bitstreams. It is the job of the encoder to calculate the drift correction signal. The drift correction signal is then coded and transmitted to the base layer decoder. The block diagrams of the encoder and the base layer decoder are shown in Fig. 6.

In the upper part of Fig. 6(a), it is a two-layer encoder using bitstream decomposition. The drift correction layer shown in the lower part of Fig. 6(a) exhibits the procedure of calculating the drift correction signal to be transmitted to the base layer decoder. The Frame Store 1 in the figure is used to store the reconstructed image with full resolution. Then encoder sends the motion vectors MVs and motion-compensated prediction image mcp1 to the drift correction layer. The so-obtained motion vectors are divided by two and used in conjunction with a bilinear interpolation filter to perform motion compensation on the previously reconstructed image, stored in Frame Store 2, with low resolution. We, then, obtain the motion-compensated prediction image mcp2 with drift error. To calculate the drift correction signal, we first perform the DCT subsampling on mcp1. That is, mcp1 is divided into 8x8 blocks, and each block is transformed by an 8x8 DCT.

Only the top left 4x4 coefficients are chosen for the next step to do 4x4 IDCT. At this time, a DCT subsampled image with quarter size of the original image is obtained. Then the difference between $mcp2$ and the DCT subsampled result of $mcp1$ is calculated. The obtained difference itself is the drift. Finally, the drift information is transformed by 4x4 DCT, quantized, zigzag scanned, and VLC coded into a separate bitstream to form the drift correction signal.

A base layer decoder with drift correction is shown in Fig. 6(b). After receiving and decoding the drift correction signal, the drift correction signal is added to the base layer to obtain reconstructed low-resolution frame. The addition of the drift correction information reduces drift and keeps the contents of the Frame Store 2 identical to that in the Frame Store of the base layer decoder.

(B) Codec-2: P1 bitstream decomposition with hybrid drift correction

In this newly proposed structure, the encoder adopts the same bitstream decomposition pattern as codec-1 and the prescribed hybrid drift correction scheme. Since the motion compensation is based on the full resolution, the motion vectors can be used directly in the drift correction layer. It is not the same as in the case of Codec-1 in which the motion vectors are divided by 2. Thus, the drift resulting from the effect of suboptimal motion vectors in Codec-1 will not happen in Codec-2. The block diagrams of the corresponding encoder and base layer decoder are shown in Fig. 7.

In the encoder design, the upper part of Fig. 7(a) is the two-layer encoder. For the drift correction layer, we first pads the chosen 4x4 low-frequency coefficients with zeros to form an 8x8 block, then performs an inverse 8x8-DCT and obtains a full-resolution image in Frame Store 2 which is reconstructed from low-frequency information only. After performing the motion compensation, DCT subsampling and 4x4 IDCT on the image stored in Frame Store 2, the predicted image $mcp2$ with minimal drift is obtained. The difference between $mcp2$ and the DCT subsampled image obtained from $mcp1$ is the drift. Again, the drift information is transformed by 4x4 DCT, quantized, zigzag scanned, and VLC coded into a separate bitstream to form the drift correction signal.

A base layer decoder with hybrid drift correction is shown in Fig. 7(b). After decoding the received drift correction signal bitstream, the decoder performs inverse quantization, pads zeros to high-frequency coefficients and executes inverse 8x8-DCT, then adding the drift correction signal to the image data decoded from the base layer. The image data after drift correction is once again stored in the Frame Store for future use.

(C) Codec-3: P2 bitstream decomposition with hybrid drift correction

In this design, as shown in Fig. 8, the same structure as codec-2 is adopted except for the bitstream decomposition

schemes and DCT subsampling methods. That is, the low-frequency coefficients of the base layer will be changed from the top left 4x4 block (in the P1 bitstream decomposition) to be the first 16 DCT coefficients in zigzag scan order (i.e. in the P2 bitstream decomposition). Moreover, the DCTs or IDCTs performed for DCT subsampling in this codec are all based on the size of 8x8, which are different for Codec-2.

V. Experimental Results

In our experiments, the video sequences Miss America and Salesman are tested. In the following, the bitrate of bitstreams is expressed in bits/frame. The objective service quality is measured by peak signal-to-noise ratio (PSNR).

Fig. 9 and Fig. 10 respectively illustrate the decoded low-resolution images with and without drift correction introduced to the base layer decoder of codec-1. Fig. 11 shows the variations of PSNR for the base layer decoder of codec-1 in the situations with and without drift correction. Furthermore, in the testing of Miss America sequence, we found that the image quality is unacceptable after five to seven frames have been coded in interframe mode, for the base layer decoder without drift correction. It is clear that the drift correction signal can significantly improve both the objective quality (expressed by PSNR), and the subjective quality (as illustrated in the snapshots).

Fig. 12 shows the variations of PSNR for the base layer decoder of codec-2 in the situations with and without drift correction. In the following, we will discuss the bitrate overhead of the drift correction and the effect the drift correction signal on codec-2. From Table 1, the bitrate overhead of codec-2 is substantially reduced as compared to codec-1. But the improved quality by adding drift correction signal to the base layer is not so obvious at this time. The reason is that the quality improvement by simply performing the drift minimization at the base layer decoder is so apparent that will mask the effectiveness of drift correction.

In videoconference applications, the corresponding motions are always small, and most of the high-frequency coefficients are zeros. That is, the top-left 4x4 block or the first 16 coefficients in zigzag scanning order of an 8x8 DCT block contains almost all the energy, and the energy contained in the high-frequency part is very small. So, the degree of quality improvement that can be achieved by drift correction is somewhat limited. In general, the base layer decoder with the ability to minimize drift effect will produce quite good service quality. The number of low-frequency coefficients included in the base layer has a great impact on the decoded base layer quality. If only a few coefficients are included in the base layer, the decoded quality will reduce and the overhead of the drift correction will increase. On the other hand, if most of the coefficients are included in the base layer, the decoded quality will increase and the overhead of the drift correction will reduce, but the real merit of the layered coding will be lost.

Table 1: The bitrate overhead of drift correction signal of codec-1 and codec-2 (Tested video sequence: Miss America).

1 I-frame+ 49 P-frame	Drift Layer (bits)
Codec-1 (Q=10)	51269
Codec-2 (Q=10)	30942
Codec-1 (Q=5)	114982
Codec-2 (Q=5)	99215

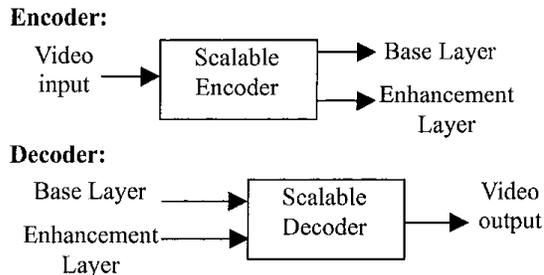


Fig. 1. Block diagram of the two-layer coding.

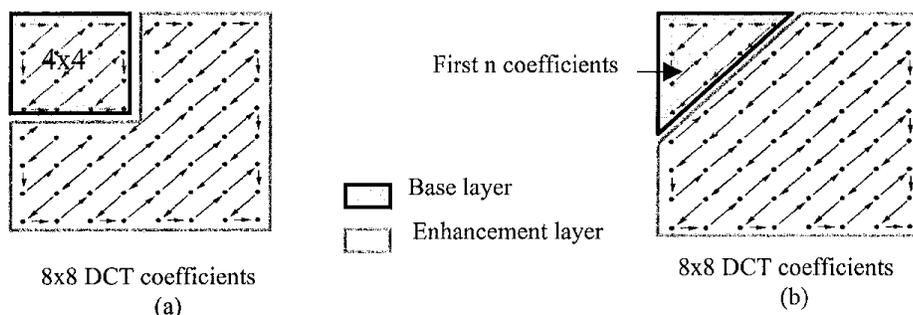


Fig. 2. The partitions of DCT coefficients for the base layer and enhancement layer. (a) Scanning pattern P1: Choosing the upper left 4x4 subblock as the base layer, and the remaining as the enhancement layer. (b) Scanning pattern P2: Choosing the first n coefficients as the base layer, and the remaining as the enhancement layer.

VI. Conclusions

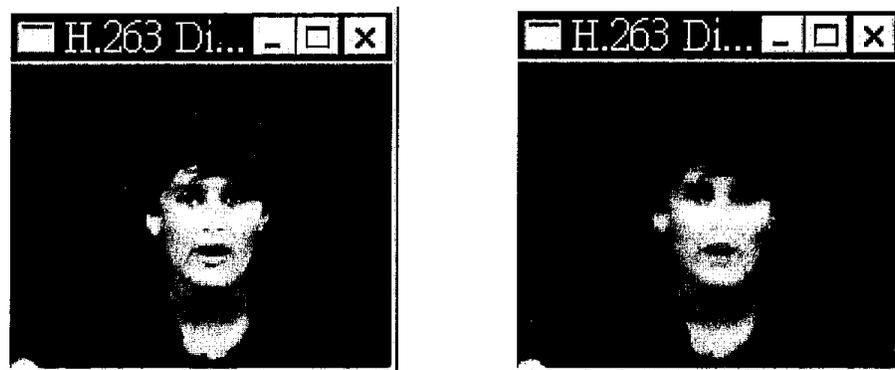
The layered coding structures described in this paper have the advantages of producing quite acceptable base layer service quality but introducing only a small bitrate overhead. To further improve the coding efficiency, we think the following two directions are worthy of doing in the future. First, one can use different VLC tables to efficiently code the base layer and the enhancement layers. Since the maximum run-length of the data contained in the base layer is 15 for the P1 bitstream decomposition, it is inappropriate to use the VLC table designated for H.263 in which the maximum run-length is 63. Second, one can design a suitable quantizer and a VLC table, to code the drift correction signal, according to its characteristics, so as to achieve better coding gain.

References

- [1] ITU-T Recommendation H.263, "Video Coding for Low Bit-Rate Communication,"
- [2] R. Mathew and J. F. Arnold, "Layered Coding Using Bitstream Decomposition with Drift Correction," *IEEE Trans. Circuits Syst. Video Technology*, vol. 7, no. 6, pp. 882-891, Dec. 1997.
- [3] R. Mokry and D. Anastassiou, "Minimal Error Drift in Frequency Scalability for Motion Compensated DCT

Coding," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 4, no. 4, pp. 392-406, Aug. 1994.

- [4] R. Aravind, M.R. Civanlar, and A. R. Reibman, "Packet Loss Resilience of MPEG-2 Scalable Video Coding Algorithms", *IEEE Trans. Circuits Syst. Video Tech.*, vol. 6, no. 5, pp. 426-435, Oct. 1996.
- [5] B. R. Lee, K. K. Park, and J. J. Hwang, "H.263-Based SNR Scalable Video Codec," *IEEE Trans. On Consumer Electronics*, vol. 43, no. 3, pp. 614-622, Aug. 1997.
- [6] M. Iwahasi, W. Kameyama, K. Ohyama, and N. Kambayashi, "A Motion Compensation Technique for Down-Scaled Pictures in Layered Coding," *IEICE Trans. Commun.*, vol. E77-B, no. 9, pp. 1007-1012, Aug. 1994.



(a)

(b)

Fig. 3. Drift effect (a) The reconstructed base layer image at the encoder without drift. (b) The decoded image obtained by the base layer decoder only.

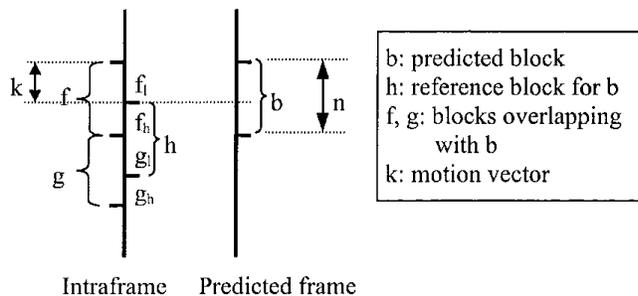


Fig. 4. The relationship between an Intra-frame (I-frame) and a predicted-frame (P-frame) for 1-D motion compensation.

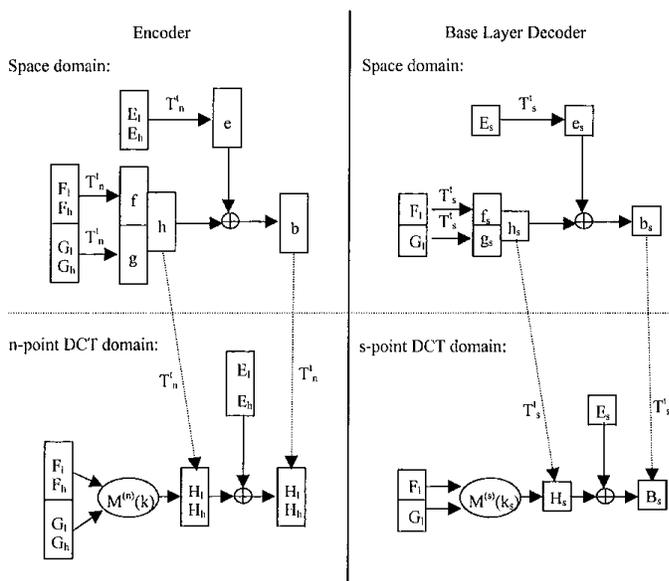
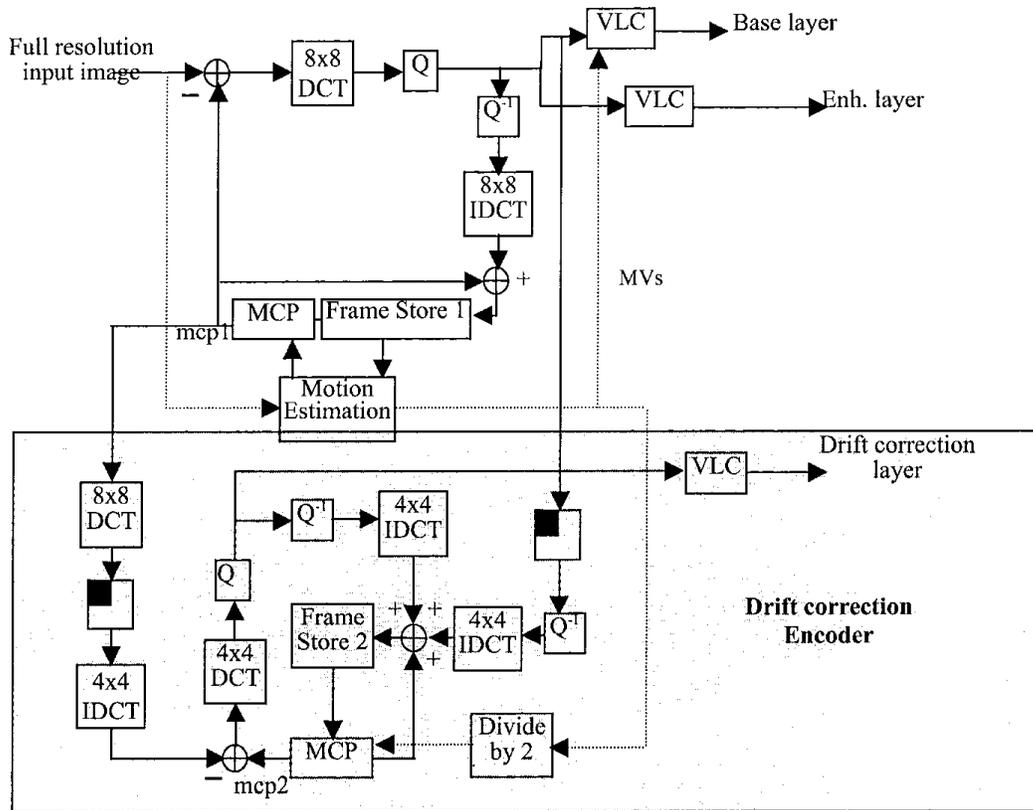
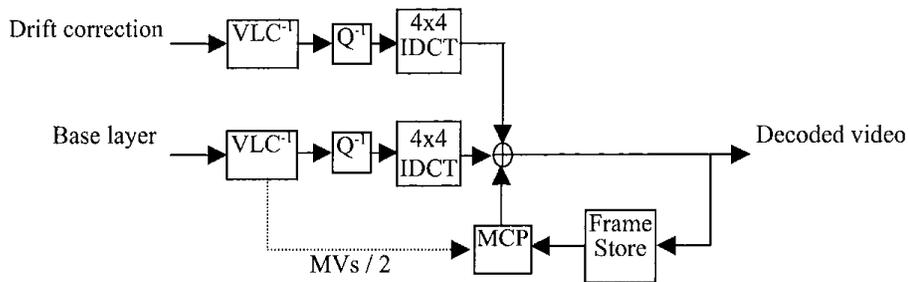


Fig. 5. The differences between the encoder and base layer decoder in both frequency domain and spatial domain, where T_n represents the n -point DCT operation, T_s represents the s -point DCT operation, and the dotted lines represent the mapping relation in spatial and frequency domains.

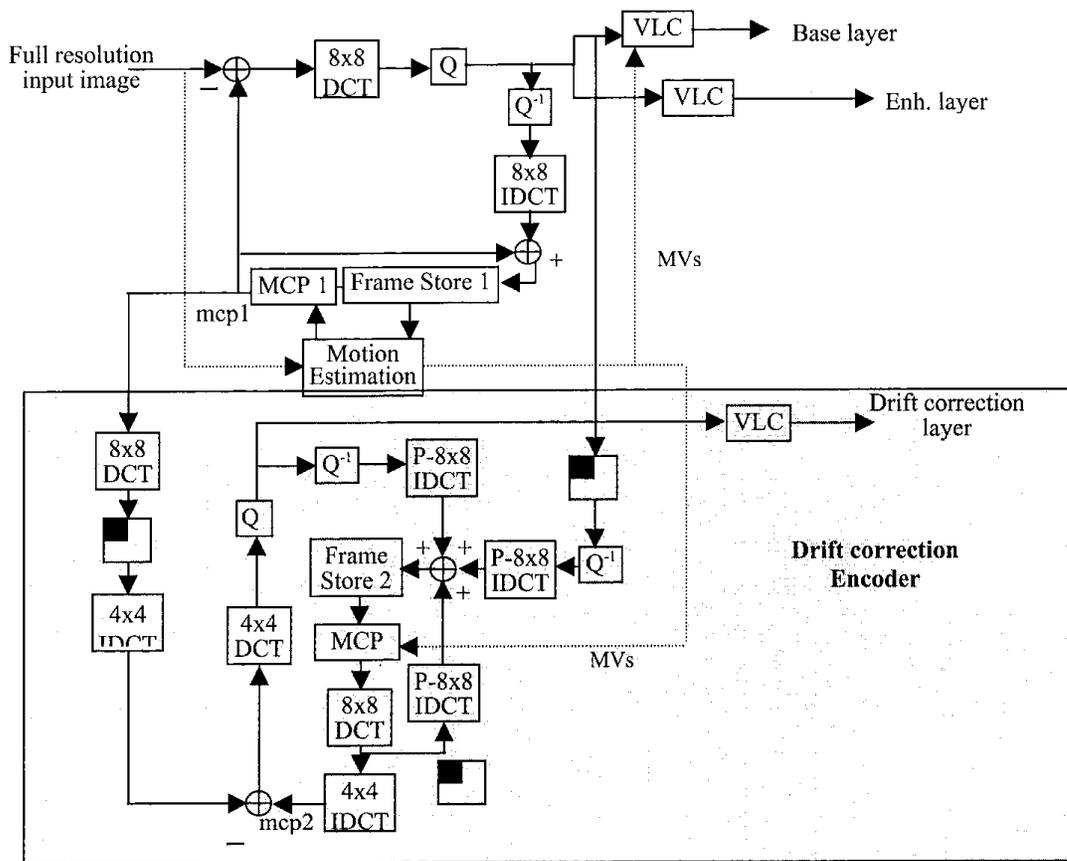


(a)

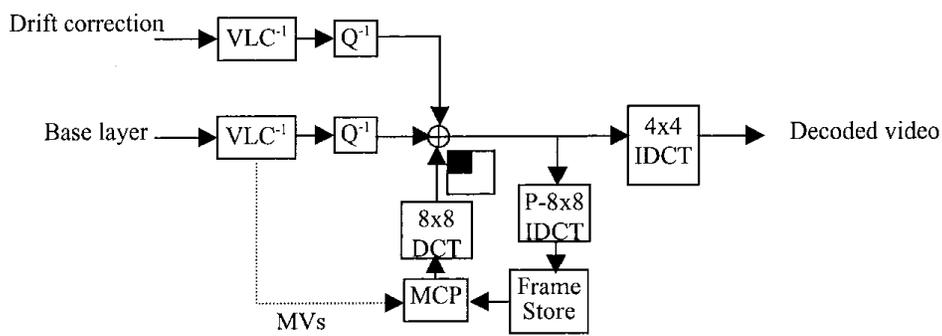


(b)

Fig. 6. Structure of Coder-1: P1 bitstream decomposition with drift correction. (a) Block diagram of encoder. (b) Block diagram of base layer decoder.

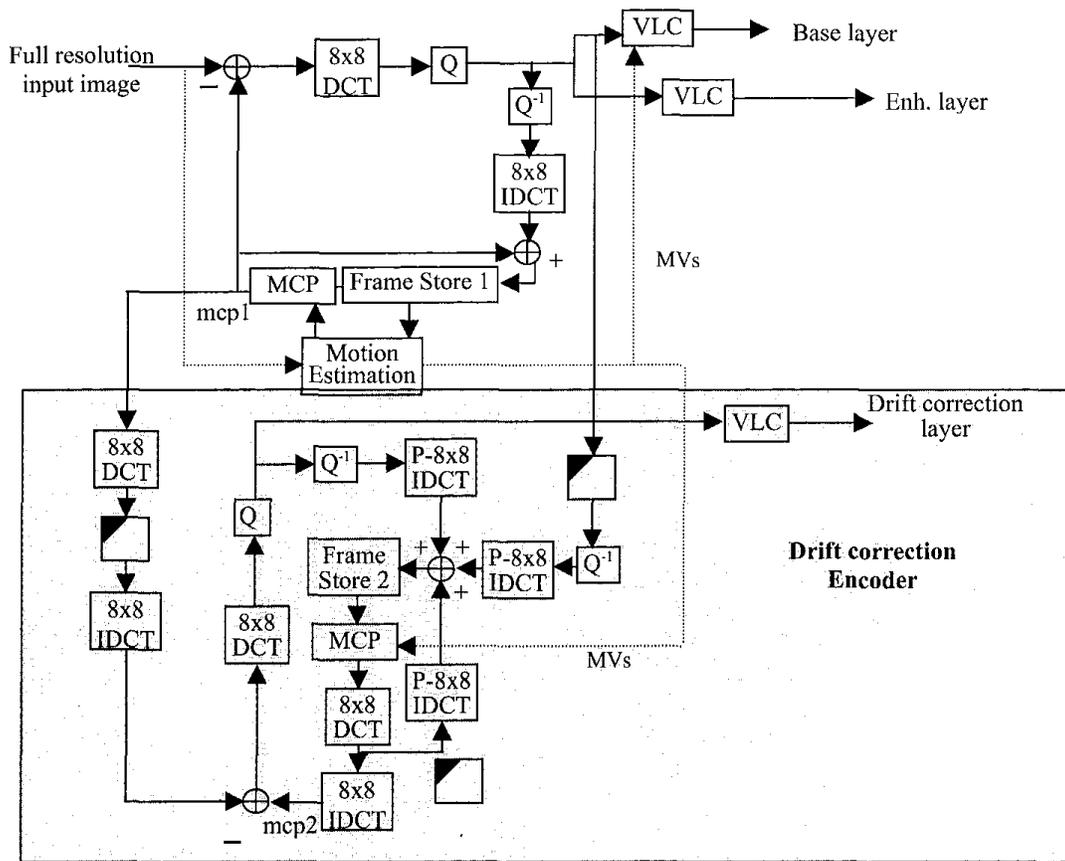


(a)

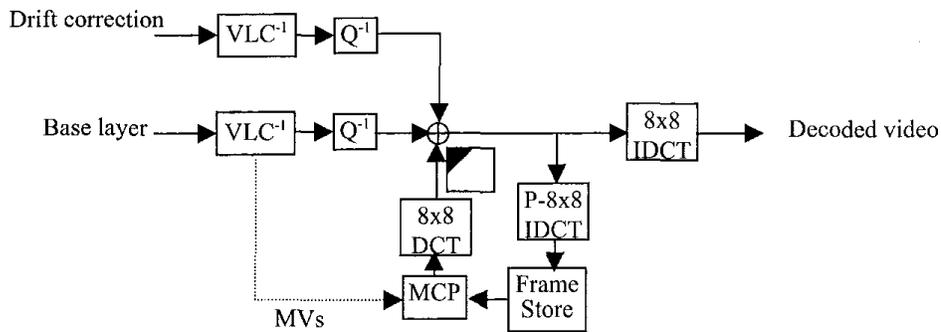


(b)

Fig. 7. Structure of Coder-2: P1 bitstream decomposition with hybrid drift correction. (a) Block diagram of encoder. (b) Block diagram of base layer decoder.



(a)



(b)

Fig. 8. Structure of Coder-3; P2 bitstream decomposition with drift correction. (a) Block diagram of encoder. (b) Block diagram of base layer decoder.

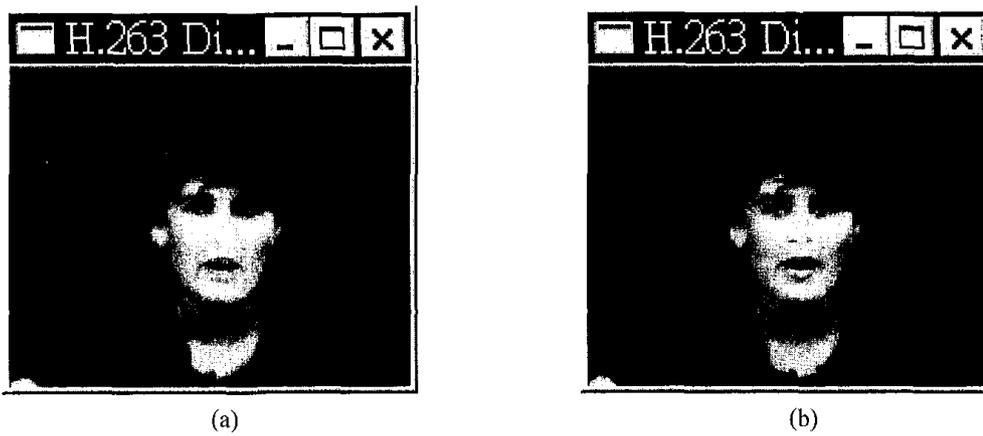


Fig. 9. (a) Frame 19 of the decoded low-resolution images with drift correction introduced to the base layer decoder (Quantizer=10) (b) Frame 19 of the decoded low-resolution images without drift correction signal. (Q = 10 used for drift correction signal).

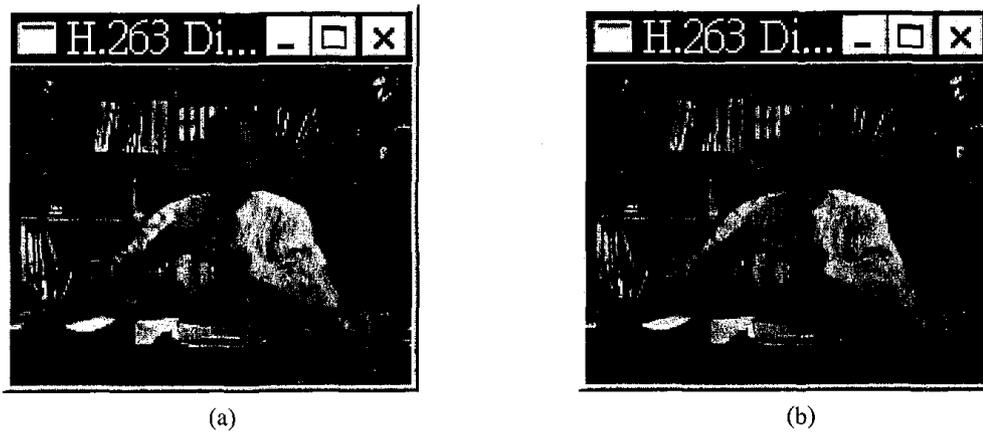


Fig. 10. (a) Frame 29 of the decoded low-resolution images with drift correction introduced for the base layer decoder (Quantizer=10). (b) Frame 29 of the decoded low-resolution images without drift correction signal (Q = 10 used for drift correction signal).

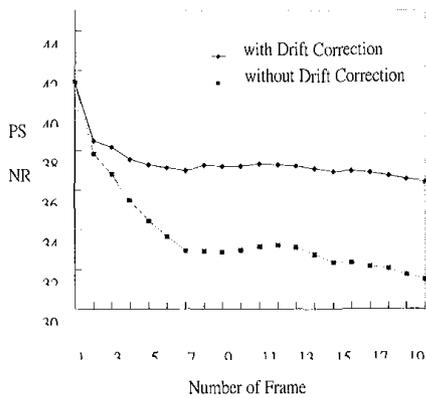


Fig. 11. Base layer PSNR versus frame number for codec-1.(Tested video sequence: Miss America)

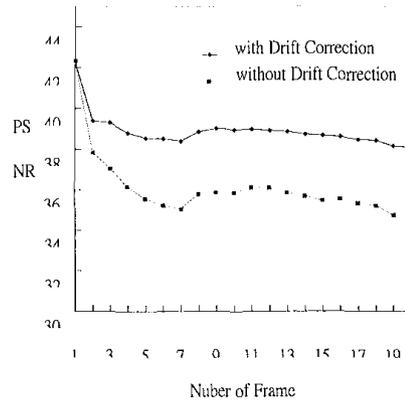


Fig. 12. Base layer PSNR versus frame number for codec-2. (Tested video sequence: Miss America)