H.264-Based Resolution, SNR and Temporal Scalable Video Transmission Systems

CHIEN-MIN OU*
Department of Electronics Engineering
Ching Yun University
Chungli, Taiwan 320, R.O.C
Taiwan

CHU-TING CHOU
Graduate Institute of Computer Science and Information Engineering
National Taiwan Normal University
Taipei, Taiwan 117, R.O.C.
Taiwan

Abstract: - This paper presents a novel H.264-based video coding scheme for a scalable delivery system which operates over heterogeneous networks and distributes real-time streaming video to diverse types of clients. The video coding scheme is a hybrid combination of discrete wavelet transform (DWT) and H.264. In the algorithm, an input video sequence is first decomposed into a fundamental sequence and a number of orthogonal supplemental sequences using DWT. Each sequence is encoded by H.264 for effective exploitation of spatial and temporal correlations. The resulting bitstreams can be organized for layered transmission, multiple description transmission, or layered multiple description transmission, depending on whether the transportation priorization is available in the network. All these transmissions provide both the SNR and resolution scalabilities. The temporal scalability can also be attained by incorporating the proposed algorithm with the motion compensated temporal filtering (MCTF) technique. Numerical results show that the proposed algorithm has superior performance over motion JPEG2000 and MPEG4. It also outperforms the H.264-based simulcast systems subject to the same transmission rate for information delivery.

Key-Words: - Video Transmission, Video Coding, Multiple Description Coding, Layered Multiple Description Coding.

1 Introduction

The demands for video streaming services have rapidly grown over the past few years. Video delivery requires a large amount of bandwidth and has high requirements on the latency and loss experienced by viewers. In addition, network environments, like Internet, are highly heterogeneous in nature. They are aggregations of connections that differ in terms of bandwidth and loss characteristics. In addition, user devices that are connected to the networks are diverse in terms of capacity and processing power. Therefore, one challenging issue in video transmission is to overcome the combined problems of the performance demands, the differences in network characteristics, and the diversity of the client devices.

One solution for flexible adaption to network and terminal capabilities is based on scalable coding, where the lower resolution or lower signal-to-noise ratio (SNR) quality video sequences are allowed to be reconstructed from partial bitstreams. Layered coding (LC) [1, 5, 14] is a technique developed for scalable video delivery, which arranges the encoded bitstreams in a hierarchical structure of accumulative layers. The layer on the bottom of the hierarchy is the base layer, and the others are the enhancement layers. The resolution or SNR quality of the reconstructed video sequences are proportional to the number of layers accumulated from the base layer by the decoder. Therefore, the base layer provides a basic level of quality and can be decoded independently of enhancement layers. On the other hand, each enhancement layer serves only to refine the layers in the lower position of the hierarchy, and alone is not useful. Hence, the base layer represents the most critical part of the scalable representation, which makes the performance of streaming services that employ layered representations sensitive to losses of base layer packets. Layered transmission therefore is

* Corresponding Author.

effective only for networks environments providing transportation prioritization.

Multiple description coding (MDC) [4, 14] has been proposed as an alternative to layered coding for streaming over unreliable networks. Each description alone can guarantee a basic level of reconstruction quality of the video source, and every additional description can further improve that quality. Although no transportation prioritization is required for multiple description transmission, each description must carry sufficient information about the original signal to guarantee an acceptable quality with a single description. This implies there will be overlap in the information contained in different descriptions. The coding efficiency therefore may be reduced as compared with the conventional single description coder.

The LC and MDC video coders based on motion compensation/estimation may also encounter drifting problem, which arises when decoders fails to reproduce high quality reference frames used by the encoder for frame prediction. In fact, the error drifting phenomena often occurs in the decoders not receiving bitstreams from all the layers or all descriptions. The scalable MPEG-2 coder [9] is a typical example, where the reference frames reconstructed only by bistream at the base layer are not identical to those used by the encoder. The MPEG-4 fine granularity scalability (FGS)[6] coder solves the drifting problem by allowing the motion compensation and prediction loop of the base layer self-contained. That is, both encoder and decoder use the same reference frames for prediction at the base layer. The residuals of the reconstructed frames are then refined at the enhancement layers. As the temporal correlation among the residuals of adjacent frames may not be high, each frame is independently encoded at the enhance layer. Consequently, substantial degradation in coding efficiency can be observed for MPEG-4 FGS as compared to its non-scalable counterpart.

This paper presents a novel drift-free scalable video eliminating the drawbacks stated above. The encoded bitstream produced by the algorithm can be adapted to LC and MDC delivery, depending on whether the transportation prioritization is available. When the algorithm is used for LC, the bitstream among different layers are non-overlapping; thereby orthogonal transmission is provided. On the other hand, when the algorithm is used for MDC, the degree of overlap among different descriptions is allowed to be pre-specified and controlled for maximizing coding efficiency while maintaining basic quality for a single description.

The algorithm is a hybrid combination of discrete wavelet transform (DWT) [12, 13] and motion compensation/ estimation. In the algorithm, a fundamental video sequences and a number of supplemental sequences are derived from the input video sequence. The fundamental sequence contains wavelet coefficients in the lowpass subband of input frames; whereas, the supplemental sequences contain the residuals of the reconstructed fundamental sequence, and the wavelet coefficients in the highpass subbands. Therefore, the fundamental sequence and supplemental sequences are disjoint. Moreover, different supplemental sequences derived from the same input sequence have disjoint sets of residuals and coefficients. This guarantees that the independent encoding the of fundamental and supplemental sequences will form an orthogonal layered transmission[7], where the base and enhancement layers contain bitstreams encoded from the fundamental and supplemental sequences, respectively.

The same bitstreams can also be used for MDC delivery, where the bitstream from each supplemental sequence will be assigned to a different description, and the bitstream from the fundamental sequence is broadcasted to all the descriptions. The amount of overlapping information among different descriptions can be contained by existing video rate control approaches[16] over the fundamental sequences with a pre-specified target rate.

A combination of LC and MDC, termed layered MDC (LMDC), is also realized using the proposed algorithm. The LMDC contains base and enhancement layers. However, unlike the LC, the enhancement layer of the LMDC is decomposed into a number of descriptions with equal importance. The LMDC has wider range of bit rates for video streaming as compared with its 2-layer LC counterparts over networks with transportation prioritization. Hence, the LMDC provides a more smooth transition in video quality by adding or deleting a description at the enhancement layer. The realization of LMDC based on our algorithm is straightforward. The bitstream encoded from fundamental sequence is assigned to the based layer; whereas, the bistream from each supplemental sequence is allocated to a different description in the enhancement layer. The descriptions in the enhancement layer are not overlapping. Accordingly, the scalable transmission is achieved with minimum overhead. The algorithms proposed in [3, 11] utilizes unequal erasure protection for attaining LMDC. The realtime transmission of the bitstreams may be difficult due to the high bandwidth overhead and long latency for channel codes delivery and decoding. On

the contrary, our algorithm requires no channel code; thereby is a low-cost solution for the realtime delivery.

Although the LC, MDC and LMDC systems realized by our algorithm achieves both the SNR and resolution scalabilities, extensions of the algorithm for temporal scalability can be accomplished by incorporating the motion compensated temporal filtering (MCTF) [2, 10] techniques. In the extensions, the fundamental and supplemental sequences are not directly derived from the input sequences. In fact, the MCTF technique is first used to decompose the input sequence into temporal lowpass and temporal highpass sequences, termed MCTF sequences. We then derive the fundamental and supplemental sequences for each of the MCTF sequences. The lowest frame rate, resolution, and fidelity can be obtained by decoding only the fundamental sequence from the lowpass MCTF sequence. The frame rate can be increased by decoding fundamental sequences from highpass MCTF sequences. In addition, the resolution and SNR can be improved by fetching the supplemental sequences from the MCTF sequences. The SNR, resolution and temporal scalabilities can therefore be attained.

In the LC, MDC and LMDC systems with or without MCTF extensions, the fundamental and supplemental sequences are encoded by the H.264 [8, 15] for efficient exploitation of temporal and spatial redundancies. The H.264 has been found to be effective for video coding by the employment of motion compensation /prediction with multiple reference frames, generalized bidirectional frames, variable block sizes, and fractional pel resolution. In addition, the adoption of H.264 also allows the existing softwares and hardwares of the standard to be reused for scalable applications. To remove the drifting problem, each sequence is independently encoded by the H.264. That is, the motion estimation /prediction loop for the encoding of each sequence is self-contained. No information from other sequences is necessary for the reconstruction of each sequence. Numerical results show that the scalable video streaming systems realized by the proposed algorithm outperform the systems implemented by MPEG-4 and motion JPEG2000. When the proposed scalable systems are also orthogonal, their performance are superior to that of the H.264-based simulcast systems subject to the same rate for information delivery.

2 Preliminaries

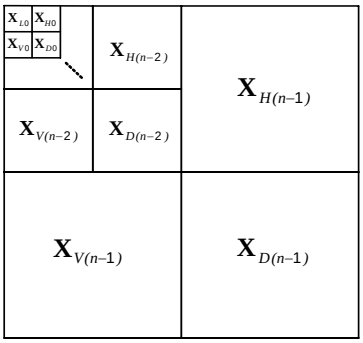


Figure 1. The DWT of a $2^n \times 2^n$ image $\mathbf{x}$.

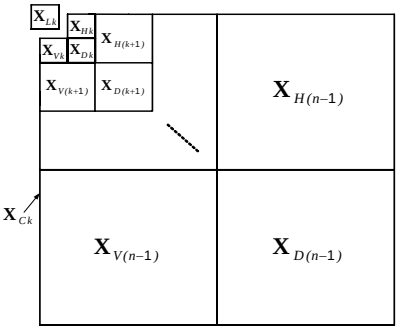


Figure 2. An example of $\mathbf{X}_{LK}$ and $\mathbf{X}_{CK}$, where the size of $\mathbf{X}_{LK}$ is $2^k \times 2^k$, and size of $\mathbf{X}_{CK}$ is $2^n \times 2^n - 2^k \times 2^k$.

This section briefly reviews some facts of the DWT, LC, and MDC. Let $\mathbf{X}$ be an image with dimension $2^n \times 2^n$. As shown in Figure 1, the DWT results of $\mathbf{X}$ contains a set of subbands $\mathbf{X}_{L0}$ an $\mathbf{X}_{Vk}$, $\mathbf{X}_{Hk}$, $\mathbf{X}_{Dk}$, $k = 0, \ldots, n-1$, each with dimension $2^k \times 2^k$. The subbands $\mathbf{X}_{Lk}$ (lowpass subbands at resolution level k), and $\mathbf{X}_{Vk}$, $\mathbf{X}_{Hk}$, $\mathbf{X}_{Dk}$, (V, H and D orientation selective highpass subbands at resolution level k), $k = 0, \ldots, n-1$, are obtained recursively from $\mathbf{X}_{L(k+1)}$ with $\mathbf{X}_{Ln} = \mathbf{X}$, where the resolution level n is also referred to as the full resolution. Conversely, the lowpass subband $\mathbf{X}_{L(k+1)}$ can be reconstructed from subbands, $\mathbf{X}_{Lk}$, $\mathbf{X}_{Vk}$, $\mathbf{X}_{Hk}$ and $\mathbf{X}_{Dk}$ by inverse DWT (IDWT). Let

$$\mathbf{X}_{Ck} = \{\mathbf{X}_{Vj}, \mathbf{X}_{Hj}, \mathbf{X}_{Dj}, j = k, \ldots, n\text{-}1\}. \quad (1)$$

An example of $\mathbf{X}_{Lk}$ and $\mathbf{X}_{Ck}$ are given in Figure 2. It is then clear that we can obtain the original image $\mathbf{X}$ from $\mathbf{X}_{Lk}$ and $\mathbf{X}_{Ck}$ by applying IDWT recursively. Both the DWT and IDWT can be carried out using a quadrature mirror filter (QMF) scheme [12, 13].

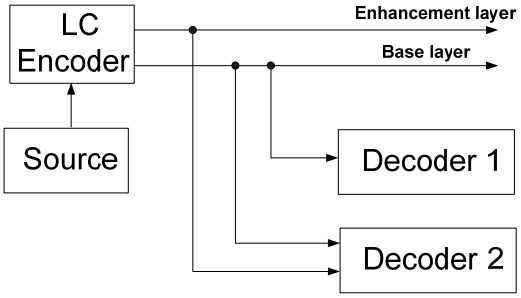


Figure 3. The basic structure of a 2-layer LC

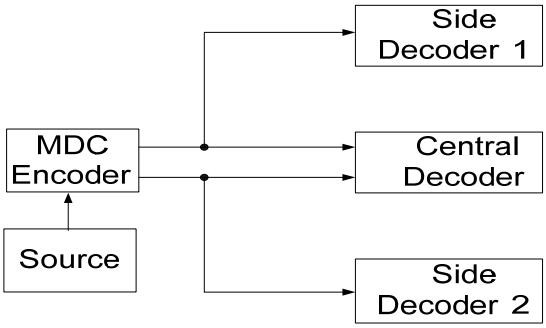


Figure 4. The basic struceure of a 2-channel MDC system.

A typical implementation of a LC system is shown in Figure 3, where the encoded bit streams for reconstructing images in different resolutions and rates are transmitted via more than one layer for decoding. Each layer is associated with a resolution level. The layers are arranged in such a way that layers having lower resolution are placed in lower positions in the system. Starting from the base layer (the layer in the lowest position, or the layer 1), the receivers can decode the bit streams *up to* any layer depending on their requirements for the reconstructed image. The resolution of the reconstructed images after decoding is the resolution of the layer in the highest position among the layers decoded by the receiver.

The MDC techniques are the effective alternatives for image/video transmission over networks without transportation prioritization. Figure 4 shows an example of a simple two-channel MDC system, where the encoded bitstreams are splitted into two channels. Each channel contains a different description of the source images. Receivers can collect bit streams from any of the two channels for frame reconstruction. In contrast to the LC schemes, where the base layer is essential for decoding, all the channels in the MDC systems have equal importance. We call the receivers receiving bitstreams from only one channel and all the channels, the side receivers and central receivers, respectively.

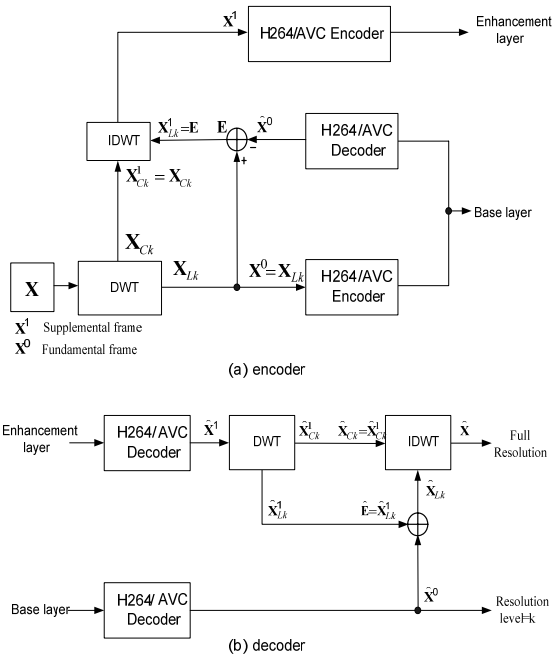


Figure 5. The proposed algorithm for the realization of a 2-layer LC system.

3 The Algorithm

This section contains three subsections. The first two subsections describe the proposed algorithm for LC and MDC designs, respectively. The extention of the algorithm for the design of LMDC systems is then presented in the final subsection.

3.1 Layered Video Transmission System

Figure 5 shows the proposed algorithm for the realization of a two-layer LC system. In the system, the resolution level associated with each layer is allowed to be pre-specified. For the sake of simplicity, assume source frames in the video sequences are of dimension $2^n \times 2^n$. They are encoded in full resolution (i.e., resolution $\text{level} = n$) at the enhancement layer, and in lower resolution (i.e., resolution $\text{level} = k$, $k < n$) at the base layer. Let $\{\mathbf{X}\}$ be the source video sequence for encoding/transmission. As shown in Figure 5, instead of compressing the input video sequence $\{\mathbf{X}\}$ directly, a fundamental sequence $\{\mathbf{X}^0\}$ and a supplemental sequence $\{\mathbf{X}^1\}$ are derived for the encoding at the base layer and enhancement layer, respectively. The derivations of a fundamental frame $\mathbf{X}^0$ and a supplemental frame $\mathbf{X}^1$ from each source frame $\mathbf{X}$ are shown as follows.

Since the resolution level associated with base layer is k , the lowpass subband $\mathbf{X}_{Lk}$ of each input

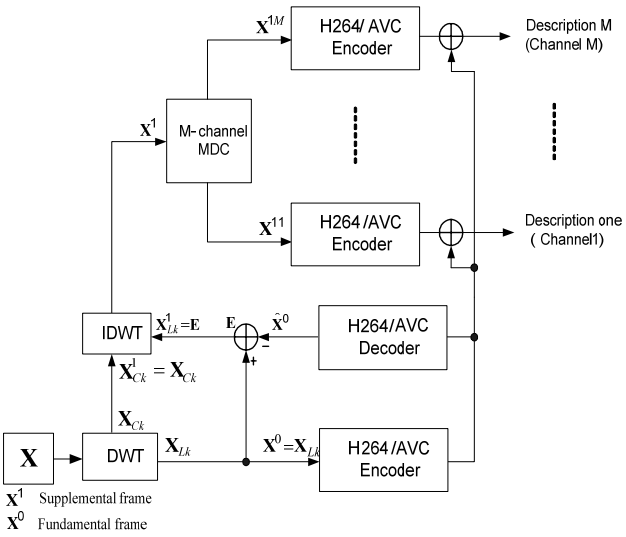


Figure 6. The proposed algorithm for the realization of the encoder of an *M*-channel MDC system.

frame $\mathbf{X}$ is used as a fundamental frame $\mathbf{X}^0$ for the encoding at the base layer. That is,

$$\mathbf{X}^0 = \mathbf{X}_{Lk}. \tag{2}$$

The encoding of $\{\mathbf{X}^0\}$ is based on H.264 for exploring the high spatial and temporal redundancies of lowpass subbands. Let $\hat{\mathbf{X}}^0$ be the reconstructed frame of $\mathbf{X}^0$, which is reproduced by the H.264 decoder in the receiver. Let

$$\mathbf{E} = \mathbf{X}^0 - \hat{\mathbf{X}}^0. \tag{3}$$

be the residual at the base layer. The $\mathbf{E}$ will be used for the subsequent encoding process at the enhancement layer for the refinement of the reconstructions at the base layer.

The enhancement layer uses the full resolution for video encoding. The supplemental frames for the encoding at the enhancement layer contain residuals at the base layer, and the highpass subbands of source frames. That is,

$$\mathbf{X}^1_{Lk} = \mathbf{E} \tag{4}$$

$$\mathbf{X}^1_{Ck} = \mathbf{X}_{Ck}. \tag{5}$$

It can then be observed the lowpass subband of $\mathbf{X}^1$ contains only the residuals of the reconstructed frames at the based layer. The subband $\mathbf{X}_{Lk}$, which has been encoded at the base layer, therefore will not be repeatedly encoded at the enhancement layer. This results in orthogonal LC transmission.

The adjacent supplemental frames $\mathbf{X}^1$'s may also be correlated because of the possible temporal redundancy exist among highpass subbands of source frames $\mathbf{X}$'s. Similar to the encoding at the base layer, this correlation can be explored by the effective motion compensation /prediction techniques of the H.264.

Let $\hat{\mathbf{X}}^1$ and $\hat{\mathbf{X}}$ be the reconstructions of $\mathbf{X}^1$ and $\mathbf{X}$, respectively. In our algorithm, the $\hat{\mathbf{X}}^1$ is also directly obtained from the H.264 decoder in the receivers. Note that, the same H.264 decoder in the receiver can be used for the reconstructions of fundamental and supplemental frames. Each reconstructed source frame $\hat{\mathbf{X}}$ is then obtained from both the $\hat{\mathbf{X}}^0$ and $\hat{\mathbf{X}}^1$ by

$$\hat{\mathbf{X}}_{Lk} = \hat{\mathbf{X}}^0 + \hat{\mathbf{X}}^1_{Lk} \tag{6}$$

$$\hat{\mathbf{X}}_{Ck} = \hat{\mathbf{X}}^1_{Ck} \tag{7}$$

Since the residual $\mathbf{E}$ is the lowpass subband of $\mathbf{X}^1_{Lk}$, the reconstructed $\mathbf{E}$, denoted by $\hat{\mathbf{E}}$, is identical to $\hat{\mathbf{X}}^1_{Lk}$. From eq. (6), it then follows that $\hat{\mathbf{X}}_{Lk}$ is the refinement of the $\hat{\mathbf{X}}^0$ at the base layer. Therefore, the bitstream of the enhancement layer provides both the reconstruction of subbands at higher resolutions and the refinement of subbands at lower resolutions.

3.2 Multiple Description Video Transmission System Based on H.264Sub-subsection

As shown in Figure 6, the same fundamental

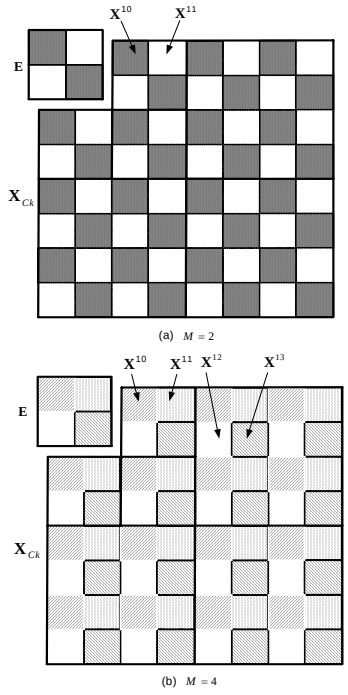


Figure 7. The partitioning of wavelet coefficients for MDC design. (a) *M*=2, (b) *M*=4.

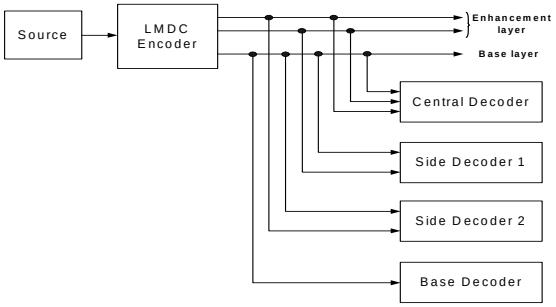


Figure 8. The basic structure of a 2-channel LMDC system.

sequence $\{\mathbf{X}^0\}$ and supplemental sequence $\{\mathbf{X}^1\}$ derived from the source sequence $\{\mathbf{X}\}$ using eqs. (2)(5) can be used for the MDC transmission. For the design of an M -channel MDC, the supplemental sequence $\{\mathbf{X}^1\}$ will be further decomposed into M orthogonal supplemental sequences $\{\mathbf{X}^{1m}\}$, $m=1,\ldots,M$. This is accomplished by partitioning the wavelet coefficients of $\{\mathbf{X}^1\}$ into M non-overlapping sets. Each set is then assigned to a different orthogonal supplemental sequence. That is,

$$\mathbf{X}^1 = \sum_{m=1}^{M} \mathbf{X}^{1m}. \tag{8}$$

Figure 7 shows one simple way for the partitioning of wavelet coefficients for $M=2$ and 4. As dipicted in figure, each subband of the frame $\mathbf{X}^1$ is first divided into non-overlapping blocks of wavelet coefficients with equal size. When $M=2$ as an example, these blocks are partitioned into 2 complementary groups (labelled either gray or white), which are then assigned to $\mathbf{X}^{10}$ and $\mathbf{X}^{11}$ as their wavelet coefficients, respectively.

All the sequences $\{\mathbf{X}^0\}$ and $\{\mathbf{X}^{1m}\}$, $m=1,\ldots,M$, are encoded separately by H.264 for efficient temporal redundancy exploitation. The bitstreams of each description m consist of the bitstreams from the fundamental sequence $\{\mathbf{X}^0\}$ and the supplemental sequence $\{\mathbf{X}^{1m}\}$, which are then delivered over the channel m of the MDC system.

Let S be the set of descriptions subscribed by a receiver. In addition, let $|S|$ be the number of descriptions in S. Therefore, $1 \le |S| < M$ for a side receiver, and $|S| = M$ for a central receiver. Let $\hat{\mathbf{X}}^{1m}$, $m=1,\ldots,M$, be the reconstructed $\mathbf{X}^{1m}$, which are obtained from the H.264 decoder in the receiver. The $\hat{\mathbf{X}}^1$ is then computed by

$$\hat{\mathbf{X}}^1 = \sum_{m\in S} \hat{\mathbf{X}}^{1m}. \tag{9}$$

From eqs. (8)(9), it then follows that the full reconstruction of $\mathbf{X}^1$ is available for a central receiver. On the other hand, side receivers do not subscribe all descriptions, and obtain only the partial reconstruction of $\mathbf{X}^1$.

The fundamental bitstream is broadcasted to all the descriptions to guarantee the basic quality of the reconstructed frames upon the receipt of a single description. Therefore, each receiver will receive up to $|S|$ identical fundamental bitstreams over packet-erasure channels. Each of the bitstreams is sufficient for the reconstruction of $\mathbf{X}^0$. Therefore, the MDC systems are less susceptible to the delivery errors of fundamental streams over networks without transportation prioritization. Although the overhead

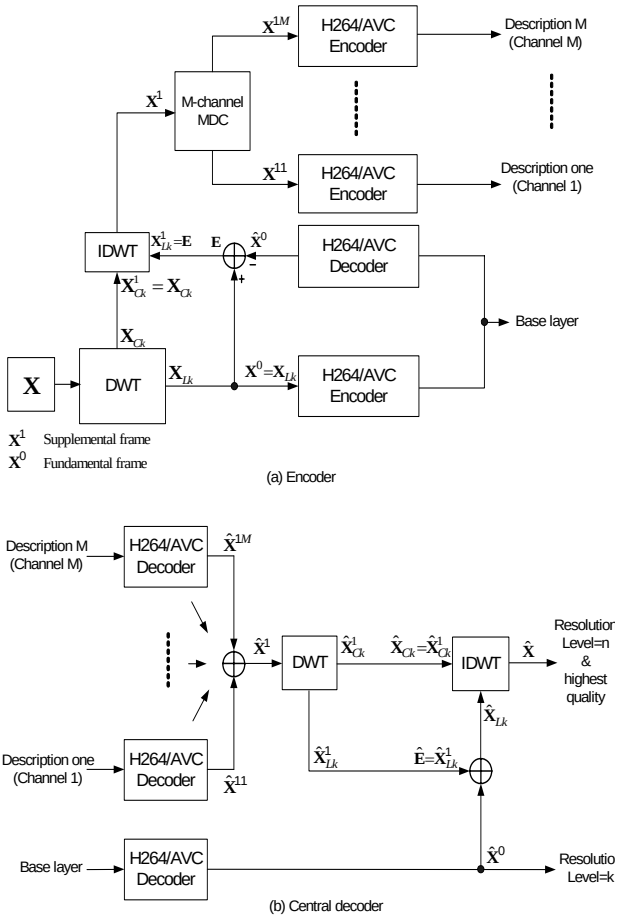


Figure 9. The proposed algorithm for the realization of an *M*-channel LMDC system. (a) Encoder; (b) Central Decoder.

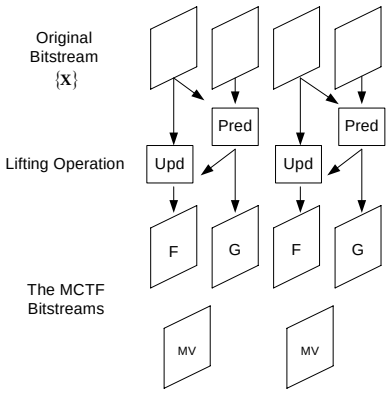


Figure 10. The basic MCTF system.

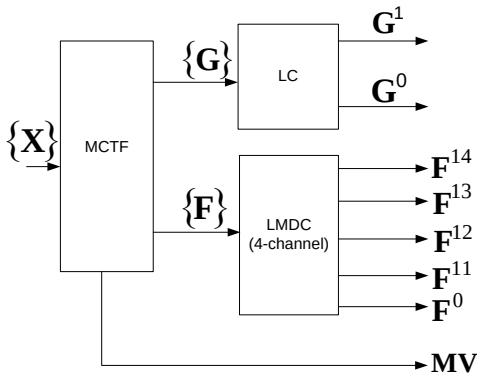


Figure 11. An example of incorporating the proposed algorithm with MCTF.

of the MDC may become substantial for large $|S|$ values, it can be contained by existing video rate control approaches [16] over the fundamental sequence with a pre-specified target rate. With a low resolution and high correlation fundamental sequence, a low target rate may be sufficient for the encoding of $\{\mathbf{X}^0\}$ by the H.264.

Finally, based on $\hat{\mathbf{X}}^0$ and $\hat{\mathbf{X}}^1$, the reconstructed source frame is obtained by the identical procedure as the LC decoder as shown in eqs. (6)(7).

3.3 Layered Multiple Description Video Transmission System

The basic structure of a basic 2-channel LMDC system is shown in Figure 8, which also contains a base layer and an enhancement layer. Similar to the LC system, the resolution level associated with base layer and enhancement layer are also given by k and $n\,(k<n)$, respectively. The enhancement layer contains two channels. Each channel consists of a different description of the supplemental video sequence. As shown in the figure, receivers have several options for the video reconstruction. They can simply subscribe bitstream from the base layer to reproduce the source video sequence in the lower SNR values and/or resolutions. To reconstruct the source video sequences with higher quality, the receivers can subscribe the bitstream from the base layer, and the bitstream from either of the channels at the enhancement layer. They can also reproduce the video sequences with highest SNR value in the full resolution by accumulating the bistreams up to both channels at the top layer.

The proposed algorithm for the implementation of the LMDC system is shown in Figure 9. The LMDC systems are also based on the fundamental sequence $\{\mathbf{X}^0\}$ and supplemental sequence $\{\mathbf{X}^1\}$ derived from the source sequence $\{\mathbf{X}\}$ using the DWT. As shown in the figure, the fundamental sequence is used for the encoding at the base layer. For an M -channel LMDC, the supplemental sequence $\{\mathbf{X}^1\}$ is adopted to derive the orthogonal supplemental sequences $\{\mathbf{X}^{1m}\}$, $m=1,\ldots,M$, identical to those in the MDC systems for the encoding at the enhancement layer. Each orthogonal sequence is served as a description of $\{\mathbf{X}^1\}$ at the enhancement layer. All the sequences are encoded independently using the H.264. The bitstream encoded from $\{\mathbf{X}^0\}$ is delivered at the base layer; whereas, the bitstream from the $\{\mathbf{X}^{1m}\}$ is transmitted over the channel m at the enhancement layer.

To reconstruct the source frames, it is necessary to receive the bitstream at the base layer, which is used to reproduce the fundamental sequence $\hat{\mathbf{X}}^0$. Note that, in the MDC, the bitstream encoded from the fundamental sequence can be obtained from any of the descriptions. However, in the LMDC, it can only be obtained from the base layer. All the descriptions contain only the supplemental bitstreams. That is, the source frames can still be reconstructed without subscribing any descriptions in the LMDC. Consequently, similar to the LC, the LMDC works well only if the delivery at the base layer is noiseless. In a receiver subscribing the bistreams at base layer, and the descriptions in a set S, the $\hat{\mathbf{X}}^1$ is also obtained using eq. (9). In addition, the eqs. (6)(7) are used for the source frame reconstruction based on $\hat{\mathbf{X}}^0$ and $\hat{\mathbf{X}}^1$. Note that, when $S=\phi$, $\hat{\mathbf{X}}^1=0$, and therefore each reconstructed source frame only contains $\hat{\mathbf{X}}^0$ as its lowpass subband. All the wavelet coefficients at the highpass subbands are identical to zero in this case.

One major advantage of the proposed algorithm is that the same encoded bitstreams of the fundamental and supplemental sequences can be adaptively configured for the MDC and LMDC delivery. The same bitstream from fundamental sequence can be delivered at the base layer of LMDC, and broadcasted to all the channels in MDC. The same bitstream from the supplemental sequence $\{\mathbf{X}^{1m}\}$ can be delivered over the channel m of the MDC, and transmitted over the channel m at the enhancement layer of LMDC. The reconfiguration of a system from LMDC to LC based on the same encoded bitstreams is also possible by aggregating the bitstreams from all the channels at the enhancement layer of LMDC into a single channel.

As compared with LC, the LMDC offers a wider range of bit rates for video streaming. In addition to the base layer, the decoder can have receptions up to M channels at the enhancement layer. For the sake of simplicity, assume each description at the enhancement layer has identical bit rate. Accordingly, there are $M+1$ bit rates available for each decoder. By contrast, the enhancement layer of the LC systems contain single channel. That is, only two possible bit rates are available for the decoders. The LMDC systems therefore are well-suited for the networks having a growing diversity of client devices. In addition, in the LC systems, subscribing or dropping the enhancement layer may result in a substantial variations in the quality of the reconstructed frames. On the contrary, the LMDC offers a smooth transition in video quality by adding or deleting a description at the enhancement layer. Our algorithm provides a flexible and effective solution to the realization of the LMDC. Therefore, it can be viewed as an effective alternative for video streaming supporting high flexibility, broad diversity, smooth transition and superior rate-distortion performance.

3.4 MCTF Extension

The LC, MDC and LMDC systems presented above achieves both SNR and resolution scalabilities. They can also be incorporated with the MCTF technique for temporal scalability extension. As shown in Figure 10, a basic MCTF scheme contains a simple lifting architecture [10] for jointly performing the temporal wavelet transform and motion compensation for each pair of input sequences to create temporal lowpass frames (denoted by $\boldsymbol{F}$), temporal highpass frames (denoted by $\boldsymbol{G}$) and motion vectors (denoted by MV). The sequence $\{\boldsymbol{F}\}$ can be viewed as the representation of the input sequence $\{\boldsymbol{X}\}$ in half frame rate. The sequence $\{\boldsymbol{F}\}$ can be further decomposed by a pyramid scheme for obtaining the input sequence representation at quarter frame rate or lower. Only one-stage decomposition is considered here for the sake of simplicity.

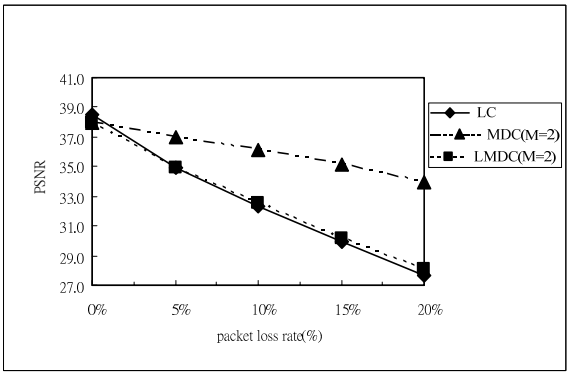


Figure 12. The average PSNR values of the LC, MDC and LMDC systems over noisy channels with various packet rates.

The combination of the proposed algorithm with MCTF can be realized by decomposing each of the $\{\boldsymbol{F}\}$ and $\{\boldsymbol{G}\}$ into fundamental sequences and supplemental sequences for LC, MDC or LMDC transmissions. By subscribing different sets of fundamental and supplemental sequences, the temporal, SNR and resolution scalabilities can be achieved. Figure 11 shows one simple example of incorporating our algorithm with the MCTF. As shown in the figure, the LMDC and LC systems are used for the delivery of encoded bitstreams of the sequences $\{\boldsymbol{F}\}$ and $\{\boldsymbol{G}\}$, respectively. The reconstructed frames with lowest frame rate, resolution and SNR can be obtained by requesting only $\{\boldsymbol{F}^0\}$. To double the frame rate while maintaining the low resolution and SNR, the

Table 1. The rate-distortion performance of the LC,MDC and LMDC systems for various source sequence.

		LC	MDC		LMDC	
			$M=2$	$M=4$	$M=2$	$M=4$
Foreman	R_T (bps)	240k	280k	360k	240k	240k
	PSNR	38.5	38.0	37.6	38.0	37.6
Carphone	R_T (bps)	240k	280k	360k	240k	240k
	PSNR	36.9	36.4	35.9	36.4	35.9
Silent	R_T (bps)	240k	280k	360k	240k	240k
	PSNR	36.4	35.9	35.5	35.9	35.5

Table 2. The rate-distortion performance of various scalable coding systems..

		LC 2-layer	LMDC		MJPEG 2k	MPEG-4	Simulcast System I	Simulcast System II
			$M=2$	$M=4$				
Football	R_T (bps)	300k	300k	300k	300k	386k	300k	400k
	PSNR	30.3	30.0	29.5	28.8	27.8	29.0	30.5
Twy	R_T (bps)	300k	300k	300k	300k	300k	300k	400k
	PSNR	39.5	38.8	38.6	35.7	37.9	38.1	39.7

receivers should then subscribe $\{\boldsymbol{F}^0\}$, $\{\boldsymbol{G}^0\}$ and the motion vectors generated by the MCTF operations. Alternatively, the receivers can decode the encoded bitstreams from $\{\boldsymbol{F}^0\}$ and some of the supplemental sequences $\{\boldsymbol{F}^{1m}\}$, $m=1,\ldots,M$, for enhancing the resolution and SNR while retaining the frame rate. Finally, the frame rate, resolution and SNR are all increased by accumulating $\{\boldsymbol{F}^0\}$, $\{\boldsymbol{G}^0\}$, some of the supplemental frames $\{\boldsymbol{G}^1\}$, $\{\boldsymbol{F}^{1m}\}$, $m=1,\ldots,M$, and the motion vectors generated by the MCTF operations. The highest frame rate, resolution and SNR quality can be attained by subscribing all the fundamental and supplemental sequences.

4 Experimental Results

This section presents some numerical results of the proposed algorithm for LC, MDC and LMDC implementations. The dimension of each frame of source video sequence is 512×512. That is, the full resolution level is $n=9$. The 5/3-tap filter [12] is used for the DWT. The resolution level of the lowpass subband $\mathbf{X}_{Lk}$ of each source frame for forming the fundamental sequences is $k=7$. Therefore, the size of each frame of the fundamental sequences is 128×128.

Let R^0 be the rate used for the encoding of the fundamental sequences. Let R^1 be the total rate used for the encoding of all the supplemental sequences. That is,

$$R^1=\sum_{m=1}^{M}R^{1m},$$

where R^{1m} denotes the rate for the encoding of $\{\mathbf{X}^{lm}\}$. Assume equal rate allocation so that

$$R^{1m}=\frac{R^1}{M}.$$

Table 1 shows the rate-distortion performance of the LC, MDC and LMDC systems for various source sequences. All the systems have identical $R^0=40\,\text{kb/sec}$, and $R^1=200\,\text{kb/sec}$. The peak SNR (PSNR) values shown in the table are defined as $10\log(255^2/D)$, where D is the mean squared distance between $\mathbf{X}$ and $\hat{\mathbf{X}}$. To compute the PSNR values, it is assumed that the decoders have receive the bitstreams encoded from all the supplemental and fundamental sequences for the full reconstruction of $\mathbf{X}$. The rate listed in the table for each system, denoted by R_T, is the rate required for the full reconstruction. Therefore, $R_T=R^0+R^1$ for both LC and LMDC systems. By contrast, $R_T=MR^0+R^1$ for the MDC systems since each description of the system contains fundamental bitstream.

From Table 1, it is observed that the LC system has slightly higher average PSNR values than the MDC and LMDC systems. This is because the sequence $\{\mathbf{X}^1\}$ is directly encoded in the LC system; whereas, $\{\mathbf{X}^1\}$ is decomposed further into sequences $\{\mathbf{X}^{1m}\}$, $m=1,\ldots,M$, before compression in the MDC and LMDC systems. Since each $\mathbf{X}^{1m}$ only holds partial information of $\mathbf{X}^1$, the intra correlation on $\mathbf{X}^1$ may not be fully exploited by the independent encoding of sequences $\{\mathbf{X}^{1m}\}$. Nevertheless, the sequences $\{\mathbf{X}^{1m}\}$ are orthogonal so that no redundancy exists among these sequences. Therefore, from Table 1, both the MDC and LMDC have marginal degradation in PSNR performance. In particular, the average PSNR value is lowered by at most 1.0 dB in these systems when $M=4$. We also note that the MDC and LMDC have the same PSNR values because they produce the same fundamental and supplemental sequences given the same R^0 and R^1.

Although the MDC has inferior rate-distortion performance as shown in Table 1, its performance is less susceptible to packet losses when the bitstreams are delivered over lossy channels without priorization. Figure 12 shows the average PSNR values of the LC, 2-channel MDC and 2-channel LMDC systems over lossy channels with various packet loss rates ε. All the specification in this experiment is identical to that in Table 1. From Figure 12, it can be observed that the performance of

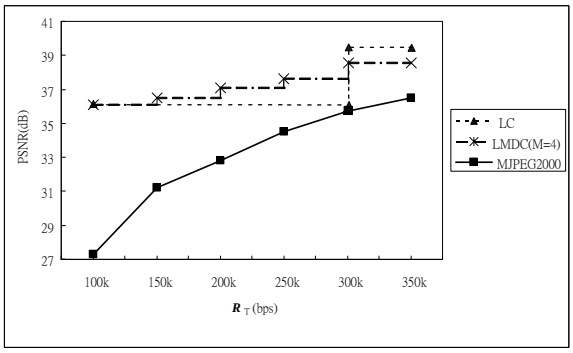


Figure 13. The transmission rate and PSNR values attainable by various scalable systems.

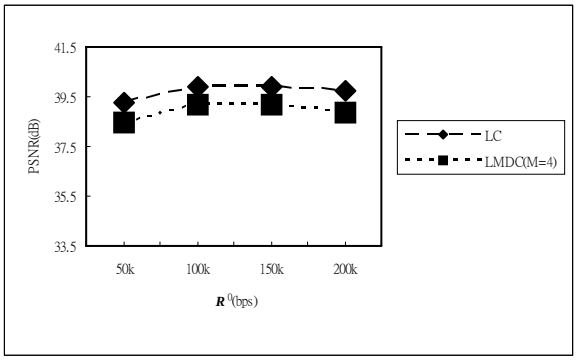


Figure 14. The PSNR values of LC and LMDC for various R^0 subject to the same total rate R_T.

the LC and LMDC systems are severely degraded as ε becomes large. By contrast, the degradation of MDC is relatively small. In particular, when $\varepsilon = 0.2$ the MDC outperforms the LC and LMDC systems by 5.9 dB and 6.4 dB, respectively.

Table 2 compares the PSNR values of various video coding algorithms measured on two 512×512 source sequences. The R^0 and R^1 for the LC and LMDC design are given by 100 kb/sec and 200 kb/sec, respectively. Therefore, both LC and LMDC have identical $R_T = 300\,\text{kb/sec}$. In these systems, the fundamental sequences are derived from the lowpass subband of the source sequences at resolution level $k = 7$. Moreover, the rate R^1 will be allocated equally to all the supplemental sequences in the LMDC. Both the motion JPEG2000 (MJPEG2000) and MPEG4 also use the same average rate 300 kb/sec for the video compression. The MJPEG2000-based systems are scalable because it produces embedded bitstreams. Although the MPEG4 is not a scalable algorithm, its rate-distortion performance can be viewed as the upper bound of that of the scalable MPEG4-FGS algorithm. From the table, we see that our LC and LMDC algorithms outperform the MJPEG2000 and MPEG4 algorithms. Our algorithms have superior performance because the H.264 technique is adopted for effective exploitation of correlation in each of the fundamental and supplemental sequences.

To further assess the performance of the LC and LMDC algorithms, the performance of two H.264-based simulcast systems are also included in Table 2. These two systems (termed Simulcast System I and Simulcast System II) attain resolution scalability by encoding the sequences $\{\mathbf{X}_{Lk}\}$ and $\{\mathbf{X}\}$ independently. In this experiment we set $k = 7$ so that $\{\mathbf{X}_{Lk}\}$ is identical to the fundamental sequence $\{\mathbf{X}^0\}$ used for the encoding of LC and LMDC systems. In both Simulcast Systems I and II, $\{\mathbf{X}_{Lk}\}$ is encoded by rate 100 kb/sec, which is also the rate R^0 used for the encoding of fundamental sequences. The rate allocated to the encoding of $\{\mathbf{X}\}$ in Simulcast System I and II are 200 kb/sec and 300 kb/sec, respectively. Therefore, the total rate for scalable coding of Simulcast System I equals to 100 + 200 = 300 kb/sec, which is the same as R_T, the total rate used by LC and LMDC. By contrast, the

Table 3 Transmission rates and their corresponding temporal, resolution and SNR qualities supported by the proposed algorithm incorporated with MCTF, where $R^0(\mathbf{F})$, $R^0(\mathbf{G})$, $R^{1m}(\mathbf{F})$, $R^1(\mathbf{G})$, and $R(\mathbf{MV})$ denote the transmission rates for the encoding of the $\{\mathbf{F}^0\}$, $\{\mathbf{G}^0\}$, $\{\mathbf{F}^{1m}\}$, $\{\mathbf{G}^1\}$ and MV, respectively.

R_T (bps)		Frame Rate	SNR	Resolution
$R^0(\mathbf{F})$	30k	half	lowest	low
$R^0(\mathbf{F}) + R^{10}(\mathbf{F})$	64k	half		full
$R^0(\mathbf{F}) + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F})$	98k	half		full
$R^0(\mathbf{F}) + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F}) + R^{12}(\mathbf{F})$	132k	half		full
$R^0(\mathbf{F}) + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F}) + R^{12}(\mathbf{F}) + R^{13}(\mathbf{F})$	166k	half	best	full
$R^0 = R^0(\mathbf{F}) + R^0(\mathbf{G}) + R^0(\mathbf{MV})$	74k	full	lowest	low
$R^0 + R^{10}(\mathbf{F})$	108k	full		full
$R^0 + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F})$	142k	full		full
$R^0 + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F}) + R^{12}(\mathbf{F})$	176k	full		full
$R^0 + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F}) + R^{12}(\mathbf{F}) + R^{13}(\mathbf{F})$	210k	full		full
$R^0 + R^{10}(\mathbf{F}) + R^{11}(\mathbf{F}) + R^{12}(\mathbf{F}) + R^{13}(\mathbf{F}) + R^1(\mathbf{G})$	240k	full	best	full

Figure 15. The original and reconstructed frames of the input sequence "carphone" of the 4-channel LMDC system with(a)original frame;(b) Transmission Rate: 40 kb/sec ; (c) Transmission Rate: 90 kb/sec ; (d) Transmission Rate: 140 kb/sec ; (e) Transmission Rate: 190 kb/sec ; (f) Transmission Rate: 240 kb/sec .

Simulcast System II has higher total rate for scalable encoding (i.e., 400 kb/sec) than R_T . It can be observed from Table 2 that both LC and LMDC has higher average PSNR values for reconstructing $\{\mathbf{X}\}$ than Simulcast System I. Moreover, using substantially lower total rate, the LC and LMDC attains comparable reconstruction fidelity to the Simulcast System II.

Although the LMDC has slightly inferior performance to LC, it offers wider range of bit rates when the full reconstruction of $\{\mathbf{X}\}$ is not necessary. Figure 13 shows the rates attainable by the LC and 4-channel LMDC systems, and their associated PSNR values for $\{\mathbf{X}\}$ reconstructions. The specification of these LC and LMDC systems is the same as that of the LC and LMDC systems

considered in Table 2. Therefore, all the systems have the same rate $R^0 = 100\,\text{kb/sec}$ for the reconstruction of fundamental sequence. The LC system has only one addition option, which is the full reconstruction of the source sequences, requiring the accumulated rate of $R^0 + R^1 = 300\,\text{kb/sec}$. By contrast, the 4-channel LMDC system has 4 additional options. The m-th option reconstructs the source sequence by acquiring the fundamental encoded bitstream, and m of the M supplemental bitstreams. The corresponding transmission rate is then given by $R^0 + m\frac{R^1}{M}$.

In the 4-channel LMDC system, the degradation in PSNR for full reconstruction as compared with the LC is only 0.9 dB. The 4-channel LMDC, however, provides 5 different rates, depending on the number of descriptions subscribed in the enhancement layer. This allows a smooth transition in video quality by adding or deleting a description at the enhancement layer, as shown in Figure 13. The performance of MJPEG2000 is also included in the figure for comparison purpose. Because the MJPEG2000 produces the embedded bitstreams, the transition in PSNR values versus rate variations is smoother than the LMDC systems. However, the MJPEG2000 has substantially lower PSNR value for full source sequence reconstruction because the algorithm does not exploit the interframe correlation.

Figure 14 shows the PSNR of 4-channel LMDC systems with different R^0 values subject to the same total rate $R_T = 300\,\text{kb/sec}$. It can be observed from the figure that only small variation in PSNR is observed for different R^0 values. In particular, the maximum variation in PSNR is only 0.7 dB for the full reconstruction of source sequence ``Foreman" as the R^0 varied from 50 kb/sec to 200 kb/sec.

Figure 15 shows the original and reconstructed frames of the source sequence “Carphone” of the 4-channel LMDC system with $R^0 = 40\,\text{kb/sec}$ and $R^1 = 200\,\text{kb/sec}$. The LMDC system offers 5 different rates: 40 kb/sec, 90 kb/sec, 140 kb/sec, 190 kb/sec, 240 kb/sec, depending on the number of supplemental streams subscribed by the decoder. Excellent visual quality is obtained by subscribing only the fundamental sequence. Graceful improvement in fidelity is also observed as the number of supplemental streams accumulated by decoder increases. Moreover, the full reconstruction has visual quality indistinguishable to that of the original frame.

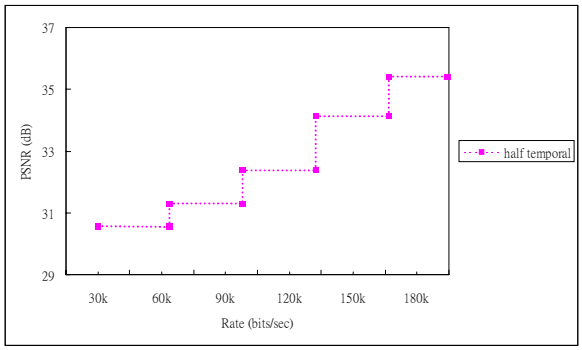
Finally, we present the performance of the system combining the proposed algorithm with the MCTF technique. The example shown in Figure 11 is realized for performance measurement, where the lifting scheme for Haar transform is used for MCTF implementation. Table 3 shows the transmission rates supported by the system and their corresponding temporal, resolution and SNR qualities. As shown in the table, there are two options for temporal scalability: half-frame rate, and full-frame rate. There are also two options for resolution scalability: low resolution and high resolution. The SNR values are dependent on the transmission rates. From the table, it can be observed that the system supports 11 different transmission rates. For the clients requiring half frame rate, only the reconstruction of $\{\mathbf{F}\}$ is necessary. Since the 4-channel LMDC (i.e., $M = 4$) is used for the encoding of $\{\mathbf{F}\}$, the system supports five transmission rates as shown in Table 3; thereby providing 5 SNR quality levels. For clients requesting full frame rate reconstruction, it is necessary to accumulate the bitstreams encoded from the $\{\mathbf{F}^0\}$ and $\{\mathbf{G}^0\}$, and some (or all) of the supplemental sequences. Six transmission rates (i.e., six SNR quality levels) are provided in this case.

Figure 16 shows the average PSNR values of the transmission rates supported by the system for the input sequence “Carphone”. Note that, since the target of reconstruction is $\{\mathbf{F}\}$ when only half frame rate is desired, the corresponding average PSNR values are measured on $\{\hat{\mathbf{F}}\}$ in the figure. On the contrary, the PSNR values are still measured on $\{\hat{\mathbf{X}}\}$ for the transmission rates supporting the full frame rate. From the figure, it can be observed that only 30 kb/sec is required when the input video sequences are delivered in lowest frame rate, resolution, and SNR quality. The average PSNR value is higher than 30 dB in this case. To reconstruct the input sequences in highest quality, the total transmission rate is then 240 kb/sec. The corresponding average PSNR is closed to 35 dB, which is comparable to average PSNR of the 4-channel LMDC system without temporal scalability. All these facts demonstrate the effectiveness of the proposed algorithm.

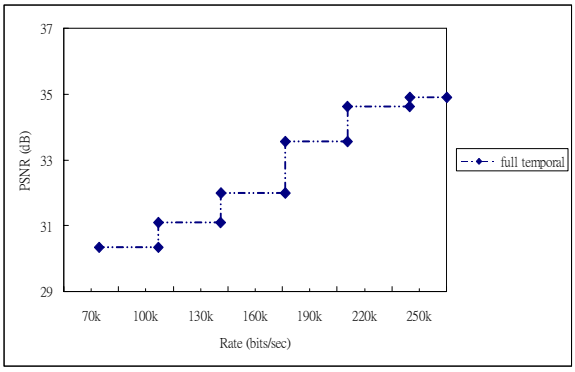
5 Conclusion

Our experiments have shown that the decomposition of source sequences into fundamental and supplemental sequences is effective for scalable video coding. The independent encoding of these sequences can be used to form the bitstreams for LC, MDC or LMDC systems. Subject to the same rate for

scalable streaming, the LC and LMDC systems have superior performance over the H.264-based simulcast systems. The LMDC systems are also able to provide graceful improvement/degradation in visual quality as the network condition varies. The performance of these systems are also insensitive to the selection of rates for fundamental bitstream delivery. In addition to providing both the resolution and SNR scalabilities, our algorithm can be combined with the MCTF technique for attaining temporal scalability. Because of its high effectiveness, flexibility and extendibility, the proposed algorithm provides an efficient tool for video streaming over heterogeneous networks.



(a)



(b)

Figure 16. The rate-distoration performance of the proposed algorithm incorporated with the MCTF. (a) The performance supporting only half frame rate; (b) The performance supporting full frame rate.

References:

[1] Bosveld F et al. Hierarchical coding, *Chap.9 in Handbook of Visual Communication*s, H.-M. Hang, J.W. Woods, Eds, Academic Press, 1995, pp.299-340.

[2] Choi S J, Wood J.W. Motion compensated 3-D subband coding of video, *IEEE Trans. Image Processing*, 1999, 8: 155-167.

[3] Chou P A *et al*. Layered multiple description coding, Packet Video Workshop, Nantes, France, April, 2003.

[4] Goyal V K. Multiple description coding: compression meets the network, *IEEE Signal Processing Magazine*, Sept., 2001, pp. 74-93.

[5] Hwang W. J. *et al.* Layered Video Coding Based on Displaced Frame Difference Prediction and Multi-resolution Block Matching, *IEEE Trans. Communications*, 2004, 1504-1513.

[6] Li W. Overview of fine granularity scalability in MPEG4 video coding standard, *IEEE Trans. Circuits and Systems for Video Technology*, 2001, 11, 301-317.

[7] Novaes M *et al.* Orthogonal layered multicast: improving the multicast transmission of multimedia streams at multiple data rates, Proc. IEEE International Conference on Communications, May 2002.

[8] Richardson I E G. H.264 and MPEG-4 video compression, John Wiley & Sons, 2003.

[9] Rao K R, Hwang J J. Techniques and standards for image, video and audio coding, Prentice Hall, 1996.

[10] Secker A, Taubman, D. Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting, Proc. IEEE International Conference on Image Processing, 2001.

[11] Stankovic V *et al.* Robust layered multiple description coding of scalable media data for multicast, *IEEE Signal Processing Letters*, 2005, 154-157.

[12] Taubman D, Marcellin M W. JPEG2000 image compression fundamentals, standards and practice, Kluwer Academic Publishers, 2002.

[13] Vetterli M, Kovacevic J. Wavelets and subband coding, Prentice Hall, 1995.

[14] Wang Y, Zhu Q F. Error control and concealment for video compression: a review, Proceedings of the IEEE, May 1998, 86, 974-997.

[15] Wiegand T *et al.* Overview of the H.264/AVC video coding standard, *IEEE Trans. Circuits and Systems for Video Technology*, 2003,13, 560-576.

[16] Xu J, He Y. A novel rate control for H.264, Proc. IEEE International Symposium on Circuits and Systems, 2004.