國立交通大學

網路工程研究所

碩 士 論 文

自動回歸模型改良之多重描述編碼

Auto-Regressive Model Enhanced Multiple Description Coding

研 究 生：溫善淳

指導教授：簡榮宏　教授

蔡文錦　教授

中 華 民 國 一 百 年 七 月

自動回歸模型改良之多重描述編碼
Auto-Regressive Model Enhanced Multiple Description Coding

研　究　生：溫善淳　　Student：Shan-Tsun Wen

指導教授：簡榮宏　　Advisor：Rong-Hong Jan

蔡文錦　　Wen-Jiin Tsai

國 立 交 通 大 學

網 路 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Network Engineering

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年七月

中文摘要

當視訊影片透過網路傳輸時，多重描述編碼是一種常被用來降低錯誤影響的技術。在本篇論文中，我們提出了一種以自動回歸模型加強的多重描述編碼。在一般的多重描述編碼當中，編碼的效能與容錯能力是評估一個方法好壞的重要標準。在我們提出的方法中使用自動回歸模型，是為了在不減低容錯能力之下，降低冗餘資訊。我們提出的多重描述編碼是由兩個對稱的描述子所組成。一個是包含有 h.264 標準的偶數幀與奇數冗餘幀，而另一個則是奇數幀與偶數冗餘幀。奇數與偶數冗餘幀都是透過自動回歸模型產生的預測幀所進一步產生出來。透過實驗證實，我們提出的方法與由內插幀預測而產生出來的冗餘幀相比，有更好的編碼效率與容錯能力。

關鍵字：多重描述編碼、自動回歸模型

ABSTRACT

Multiple description video coding (MDC) [1] is one of popular solutions to reduce the detrimental effects caused by transmission over error-prone networks. In this thesis, an auto-regressive model enhanced MDC is proposed. In general MDC architecture, redundancy rate and error resilience performance are important criterion for assessment. Auto-regressive model adopted in our proposal aims at reducing the redundancy rate while keeping the error resilience performance in our proposal. The proposed MDC model comprises two symmetric descriptions. One description is composed of even frames in h.264 standard and odd residual frames; while the other is composed of odd frames and even residual frames. Both even and odd residual frames use the prediction frames generated by auto-regressive model. The experiments show that it achieves better coding efficiency and error resilience than descriptions which residual frames are predicted from interpolated frames [2] in packet loss networks.

Index Terms----Multiple Description Coding, Auto-regressive Model

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Multimedia applications and services develop rapidly with the network progressing. However, these applications and services suffer from the unstable network quality, such as packet loss and variation of packet loss. When compressed video is transmitted over such network, it may be decoded incorrectly and decrease the coding quality. A popular solution to this problem is Multiple Description Coding (MDC), which offers a promising way to improve error resilience over error prone network. Nevertheless, MDC introduces redundancy to guarantee the quality at the decoding side. The more redundancy, the better error resilience for a MDC can have.

There are many different approaches for MDC architecture. Due to simplicity, a number of schemes are based on information splitting. In [3], the input sequences is split in temporal domain, while [4] in spatial domain and [5] in frequency domain.

S. Adedoyin *et al*. [2] proposed a scalable MDC with side information using motion interpolation (MCI-R). It splits the original sequences into two sub-sequences, one for even frames, and odd frames for the other. Each sub-sequence is encoded into one description of primary frames of its own and redundant frames which is the frames from the other sub-sequence. For example, one description comprises primary even frames and redundant odd frames, vice versa. The redundant frames are predicted from interpolated frames by primary reconstructed frames. If primary frame is corrupted during decoding, taking frame T as an example, frame T-1 and T+1 can first interpolated a predicted interpolated frame TI, and by redundant version of T, it improves TI to T', which is better in quality than TI.

Auto-regressive model has been widely applied in image processing. In this paper, we further improved [2] by enhancing its interpolated frame using auto-regressive model. Yongbing Zhang *et al*. [6] proposed a spatial-temporal auto-regressive model (STAR) for frame rate up-conversion (FRUC). Each pixel in STAR is modeled as a linear combination of spatial and temporal neighboring pixels. The weighting coefficients of the neighboring pixels are derived from an iterative self-feedback weighting training algorithm. It can further handle some complicated situations such as zooming and panning. We revised STAR to temporal auto-regressive model (TAR) and employed it into our multiple description coding architecture. Each Pixel in TAR is modeled as a linear combination of only temporal neighboring pixels. Redundant frame is predicted from the frame generated by auto-regressive model (AR frame) in our proposal. After prediction, residual, the differences between AR frame and the original frame, is calculated and redundant frame is encoded by doing DCT and Quantization from the residual.

TAR model is also applied into error concealment strategy in our proposal. Hence, it gets better quality even if the side information is lost. By employing redundant frames based on TAR model, our proposed MDC approach performs well under lossy network when the redundant rate is lower than half of primary rate.

In later section, a more detailed introduction of the related work is presented followed by our motivation. In section 3, auto-regressive model enhanced multiple description coding, our proposed method, is introduced. We provided experimental results in section 4 and summarized conclusion at the final section.

Chapter 2 Related Works

2.1 Multiple Description Coding

In multiple description coding (MDC) [1], original encoding video stream is coded into at least two equally important sub-streams. Each sub-stream, which is also called a description, can be decoded independently. With more descriptions, quality of decoded streams will get better. In comparison of single description coding (SDC), MDC descriptions are sent in different channels while SDC sends entire stream in only one channel. Hence, MDC architecture takes advantages on less probability of losing all descriptions under the assumption that packet loss rates in different channels are identical independent distributed.

An important issue of designing MDC architecture is to improve decoding quality of each description and lower the total bit-rate of all descriptions. Figure 2-1 is a typical MDC structure. The input source is encoded into two descriptions, and each description should be sent through different channels. At decoder side, if only one descriptor is received, the side decoder will decode the one-description bit-stream. In our example, which side decoder is used depends on which description is received. If both descriptions are received, the center decoder is responsible to decode and generate output in the best quality.

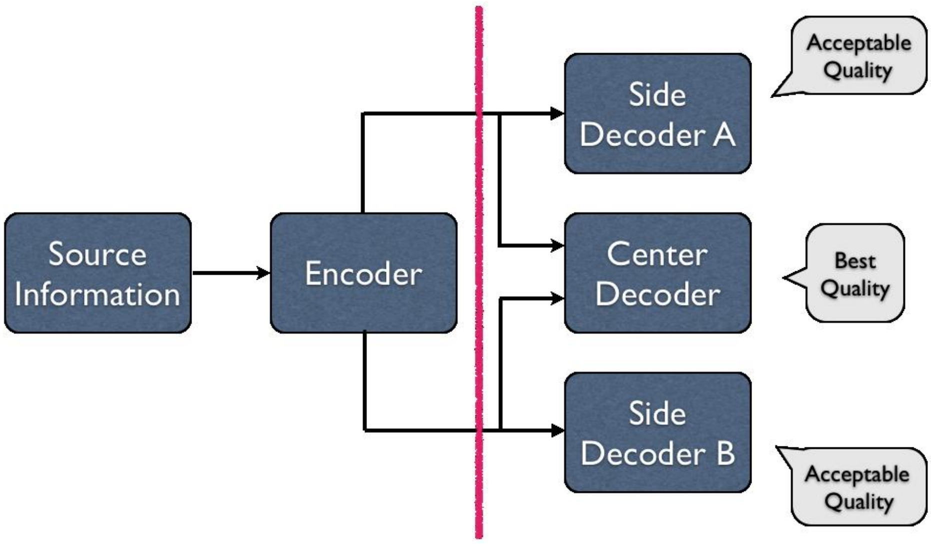


Figure 2-1 Conventional MDC Structure

2.2 Scalable Multiple Description Coding with Side Information Using Motion Interpolation

In order to acquire more error resilience for transmitting video over network, most MDC architecture is designed to carry more information apart from original sequences. In [7], multiple description scalar quantizer (MDSQ), the first MD coder, is proposed by Vaishampayan. It mapped quantized coefficients to two indices of a index assignment table. The redundant information in MDSQ is the coefficient. Redundant slice is used to enhance MDC in [8].

S. Adedoyin *et al*. proposed a scalable MDC with side information using motion interpolation (MCI-R). The redundant information in its MDC architecture is residual coded interpolating frame. Figure 2-2 shows the basic structure of descriptions of MCI-R. Based on even and odd frames structure, each description carries either even frames or odd frames as its primary data and remaining frames as redundant frames.

Thus, Primary frames and redundant frames are arranged interleaved.

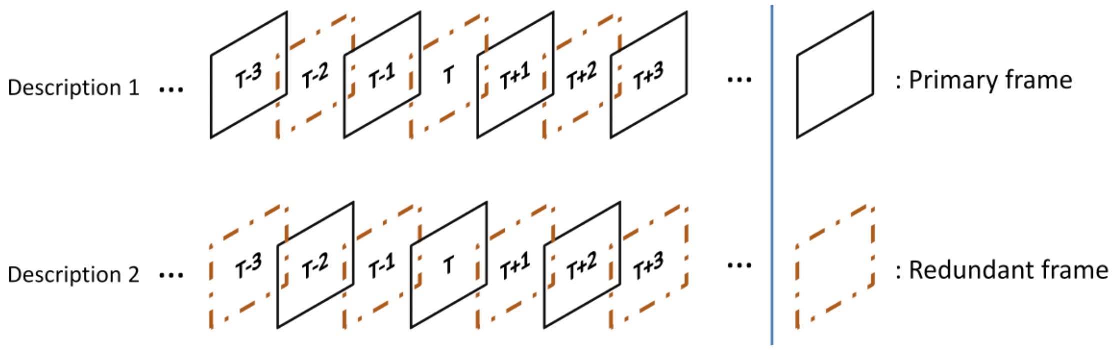


Figure 2-2 MCI-R Description Structure

During the encoding procedure, primary frames are predicted from the previous reconstructed primary frame. Take description 2 in Figure 2-2 as an example, primary frame T is predicted from reconstructed frame T-2. To encode redundant frames, the next primary frame should be encoded first. As the redundant frame T in description 1 in Figure 2-2, frame T+1 should be encoded first. According to reconstructed frame T-1 and frame T+1, these two frames may interpolate an intermediate frame T, which we called an interpolated frame. The interpolated frame T is then used as a prediction for encoding frame T. With the predicted frame, interpolated frame as well, the residual information of frame T is calculated by subtracting the interpolated frame from the corresponding frame in other description, which should be a primary frame. The residual information is then embedded back to stream as a redundant frame after residual coding, which consists of discrete cosine transform (DCT), Quantization (Q) and entropy coding.

The decoding procedure of MCI-R is as follows. First, if the to-be-decoded primary frame is received correctly, decode it directly. If the to-be-decoded primary frame is corrupted, the error concealment strategy is activated depending on the status of next primary frame. When the next primary frame is received, motion interpolation is performed with residual frame improving the motion interpolated frame. Otherwise, there is consecutive loss of primary frames, and a whole frame loss error concealment strategy such as frame copy is used. Since each description is independently coded, when only one single description is received at the decoder, it can still be decoded with good quality. It also takes advantage of having no mismatch occurring at the decoder.

2.3 STAR model

In order to improve the visual quality of interpolated frames, Spatial-Temporal auto-regressive model (STAR) is proposed in [6] by Zhang *et al*. Each pixel is modeled as a linear combination of its temporal and spatial neighborhood. The frames are first divided into non-overlapping area with size $W_x x W_y$, said training window *R*. By the model, it assumes each pixel in a training window is interpolated by corresponding spatial-temporal neighborhoods using the same weighting vector $\vec{w}$. With the aid of least square method, the best fitting weighting vector is solved.

Figure 2-3 shows the STAR model diagram. Each pixel in the to-be-interpolated training window is formulated as (1) where $\widehat{R}_{t-1}$ is the to-be-interpolated pixel; W_p, W_f, W_s represent the weights of temporal neighborhood in the previous frame, the weights of temporal neighborhood in the following frame, and the weights of spatial neighborhood, respectively.

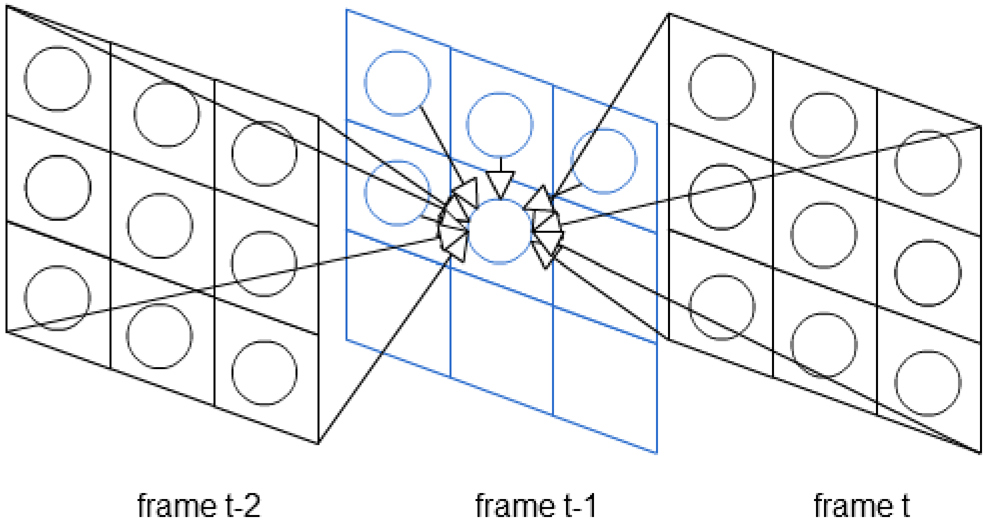


Figure 2-3 STAR Model Diagram

$$\widehat{R}_{t-1}(k,l) = \sum_{-L\leq(u,v)\leq L}\sum R_{t-2}(k+u,l+v)\times W_p(u,v) + \sum_{-L\leq(u,v)\leq L}\sum R_t(k+u,l+v)\times W_f(u,v)$$
$$+ \sum_{\{v<0,-L\leq u\leq L\}\cup\{v=0,-L\leq u<0\}}\sum \widehat{R}_{t-1}(k+u,l+v)\times W_s(u,v)$$

(1)

Spatial-temporal support order is defined as L. When L is set to 1, the pixel is modeled as the weighted sum of 9 pixels in previous frame, 9 pixels in following frame, and 4 pixels in current frame. The (k, l) represents the pixel location within the training window. The (u,v) represents looping index for each element in spatial-temporal neighborhood, which is called support region. The optimal solution for weighting vector is the one that minimizes the distortion ε between training windows $\widehat{R}_{t-1}$ and R_{t-1} where ε is defined in equation (2).

$$\varepsilon = E\left(R_{t-1}-\widehat{R}_{t-1}\right) = \sum_{k=0}^{Wx}\sum_{l=0}^{Wy} E\left[\left(R_{t-1}(k,l)-\widehat{R}_{t-1}(k,l)\right)^2\right]$$

(2)

As the actual pixel values in the interpolated frame are unknown, the estimation of ε is not feasible. Therefore, self-feedback weight training loop algorithm is introduced to solve the problem. It is an iterative method which is composed of two stages. First, the pixels in training windows $\widehat{R}_{t-1}$ and $\widehat{R}_{t+1}$ are interpolated according to their spatial-temporal neighborhood with weighting vector $\vec{w}$ where $\vec{w}$ is composed of the elements of W_p, W_f, and W_s in manner of one dimension.

Second, the training window $\widehat{R}_t$ is approximated by the pixels of $\widehat{R}_{t-1}$ and $\widehat{R}_{t+1}$ with the same weighting vector $\vec{w}$, as in Figure 2-4 and the equation (3).

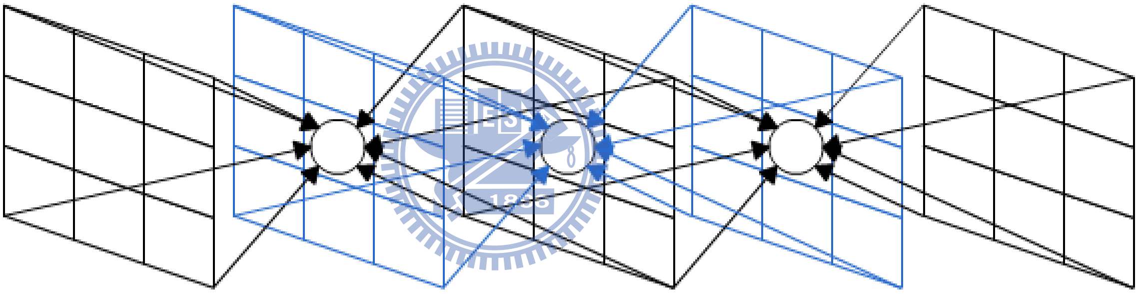


Figure 2-4 Self-feedback algorithm diagram

$$\widehat{R}_t(k,l) = \sum_{-L\leq(u,v)\leq L}\sum \widehat{R}_{t-1}(k+u,l+v)\times W_p(u,v) + \sum_{-L\leq(u,v)\leq L}\sum \widehat{R}_{t+1}(k+u,l+v)\times W_f(u,v)$$

$$+\sum_{\{v<0,-L\leq u\leq L\}\cup\{v=0,-L\leq u<0\}}\sum \widehat{R}_t(k+u,l+v)\times W_s(u,v) \qquad \textbf{(3)}$$

After all of $\widehat{R}_{t-1}$, $\widehat{R}_t$, $\widehat{R}_{t+1}$ have been interpolated, the joint distortion is calculated as equation (4).

$$D(i) = \sum_{k=0}^{W_x}\sum_{l=0}^{W_y} E\left[\left(\hat{R}_{t-1}^{i+1}(k,l) - \hat{R}_{t-1}^{i}(k,l)\right)^2\right] + \sum_{k=0}^{W_x}\sum_{l=0}^{W_y} E\left[\left(\hat{R}_{t+1}^{i+1}(k,l) - \hat{R}_{t+1}^{i}(k,l)\right)^2\right] + \sum_{k=0}^{W_x}\sum_{l=0}^{W_y} E\left[\left(\hat{R}_{t}^{i+1}(k,l) - R_t(k,l)\right)^2\right] \tag{4}$$

In the equation, *i*. represents the iteration index, while $\hat{R}_{t-1}^{i}$ and $\hat{R}_{t-1}^{i+1}$ are the interpolated training windows prior to and after the *i*th iteration respectively. By Linear least square method, the joint distortion $D(i)$ is minimized and the actual weighting vector is derived. The weighting vector after *i*th iteration is then be defined as (5) in manner of one dimension.

$$\vec{W}^i = [\vec{W}_p^i, \vec{W}_f^i, \vec{W}_s^i]^T \tag{5}$$

As long as the training windows $\hat{R}_{t-1}^{i}$ and $\hat{R}_{t+1}^{i}$ prior to the *i*th iteration are available, the weighting vector $\vec{W}^i$ can be computed by the closed form of the least square method as (6)

$$\vec{W}^i = \left({A^i}^T A^i\right)^{-1} {A^i}^T \vec{b}^i \tag{6}$$

Here the A^i in the equation is a matrix, while $\vec{b}^i$ is a column vector, and T stands for the transpose operation for a matrix. According to the regression model, A^i is defined by the regressors and $\vec{b}^i$ is defined by the regressands.

The summarization of STAR model and the self-feedback weight training algorithm is as follows:

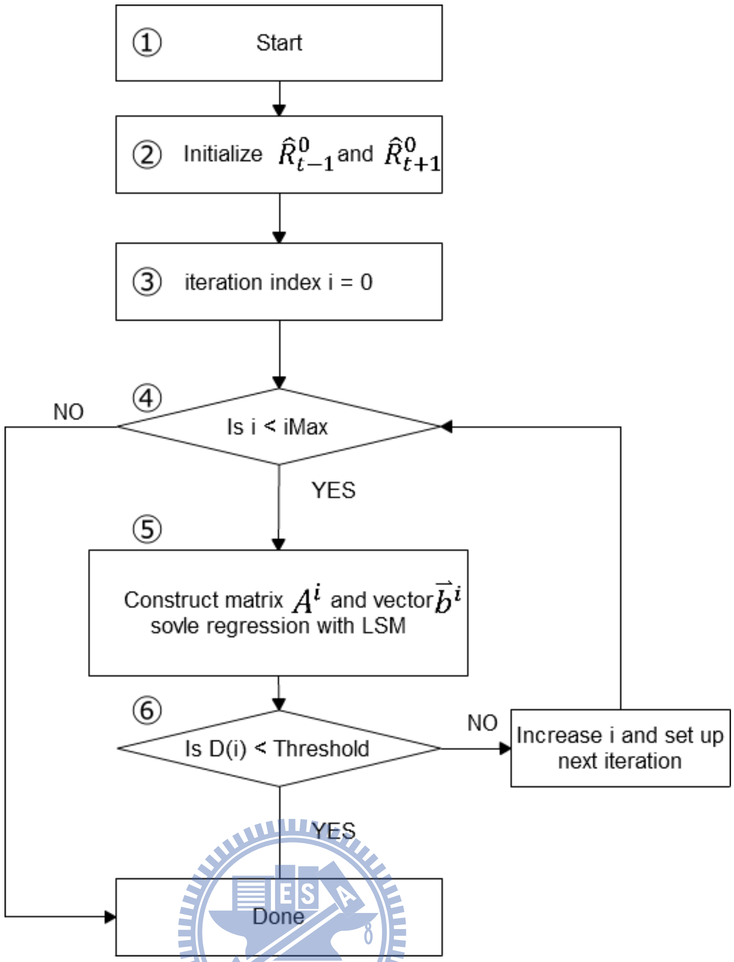


Figure 2-5 Flow chart of STAR model and self-feedback weight training

Step 1: Setting up model parameters, such as training window size W_x, W_y, jointly distortion threshold, maximum iteration times (iMAX) …etc.

Step 2: Use motion compensated interpolation frame as its initial value for training windows $\hat{R}^0_{t-1}$ and $\hat{R}^0_{t+1}$

Step 3: Initialize iteration index

Step 4 and 5: Construct corresponding matrices and vectors for regression. After least square method, D(i) of equation (4) is calculated.

Step 6: If D(i) is less than the predefined threshold, the procedure is done. Otherwise, increase the iteration index, write back the result of the new training windows as the initial value of next iteration, and loop again.

Chapter 3 Proposed Method

3.1 Motivation

Due to redundant information such as low resolution frame, frame with larger QP and other approaches, more error resilience is gained for MDC architecture depending on what it carries besides original sequences. In [2], interpolated frames are employed to lower redundancy rate and to improve distortion when the primary frame is lost. In order to further reduce redundancy rate, we were looking for another way to generate interpolated frame. With better interpolated frame, the redundancy rate descends with better prediction. Hence, we employed auto-regressive model in our proposal.

After massive observation on the weighting behavior of STAR model, we found the spatial support region can be ignored in some cases because the information of spatial region is usually included in temporal region especially when the temporal support region is higher. Thus we revised STAR model to TAR model, which the spatial support region in regression model is neglected. By embedding TAR model into our MDC structure, we improved [2] on its interpolated frame to achieve better performance in rate and distortion.

3.2 Encoder

Our proposed MDC is based on the even and odd MDC structure. Figure 3-1 illustrates the proposed encoding architecture.

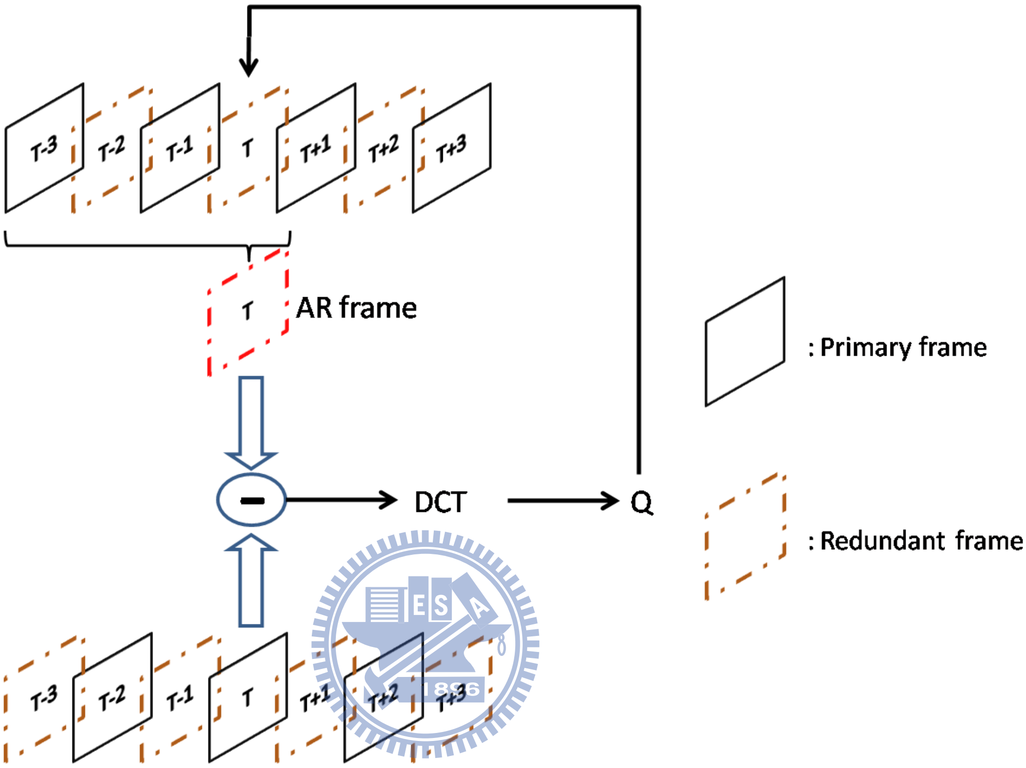


Figure 3-1 Proposed MDC with Even-Odd Frame and Residual Encoder

First we split the original sequences into two sub-sequences, and generate two descriptors. One descriptor is including primary even frames and redundant residual odd frames, and the other descriptor is including primary odd frames and redundant residual even frames. To encode a primary frame, it is predicted from previous primary frame. Taking Figure 3-1 as an example, primary frame T+1 is predicted from primary frame T-1. To encode a redundant frame T, the residual information is generated by comparing the difference between the auto-regressive frame (AR frame) and the reconstructed primary frame of the other side encoder. To obtain the AR frame of frame T, the next primary frame, frame T+1, should be encoded first. With the consecutive three reconstructed primary frames T-3, T-1, and T+1, AR frame T is interpolated by TAR model. After the difference is computed, the redundant frame is generated by residual coding including DCT, quantization, and entropy coding and then it is embedded back into the stream. According to the scheme we use, it is allowed for user to control the redundancy rate by setting the quantization parameter. The redundant frame is not used as reference frame for primary frames.

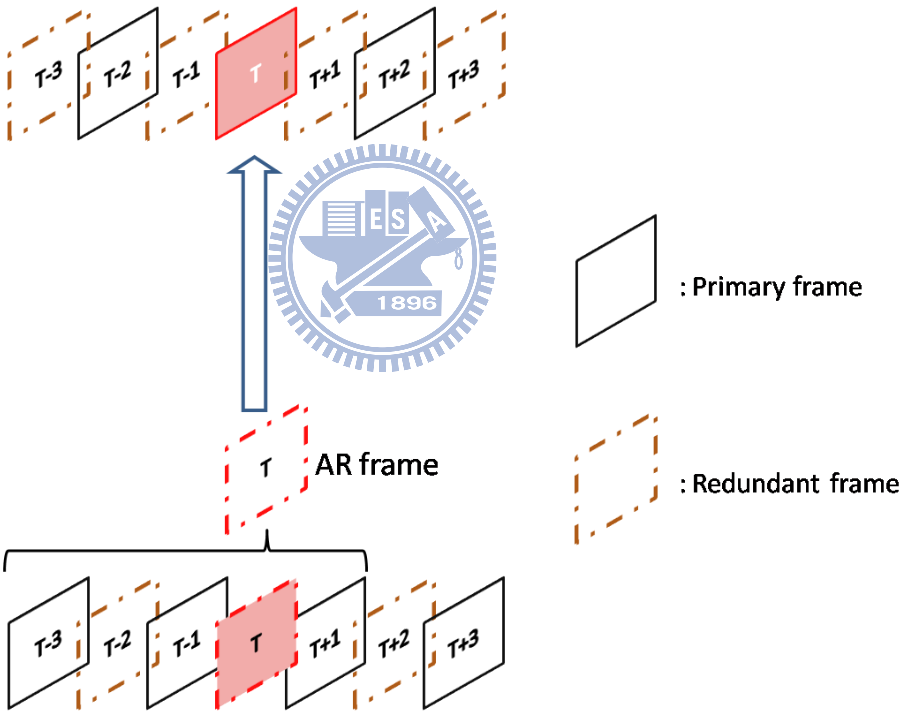
3.3 Decoder

At the decoder side of our model, there are different strategies switching according to the packet loss environment. Table 3-1 shows the strategy it takes under different packet loss circumstances.

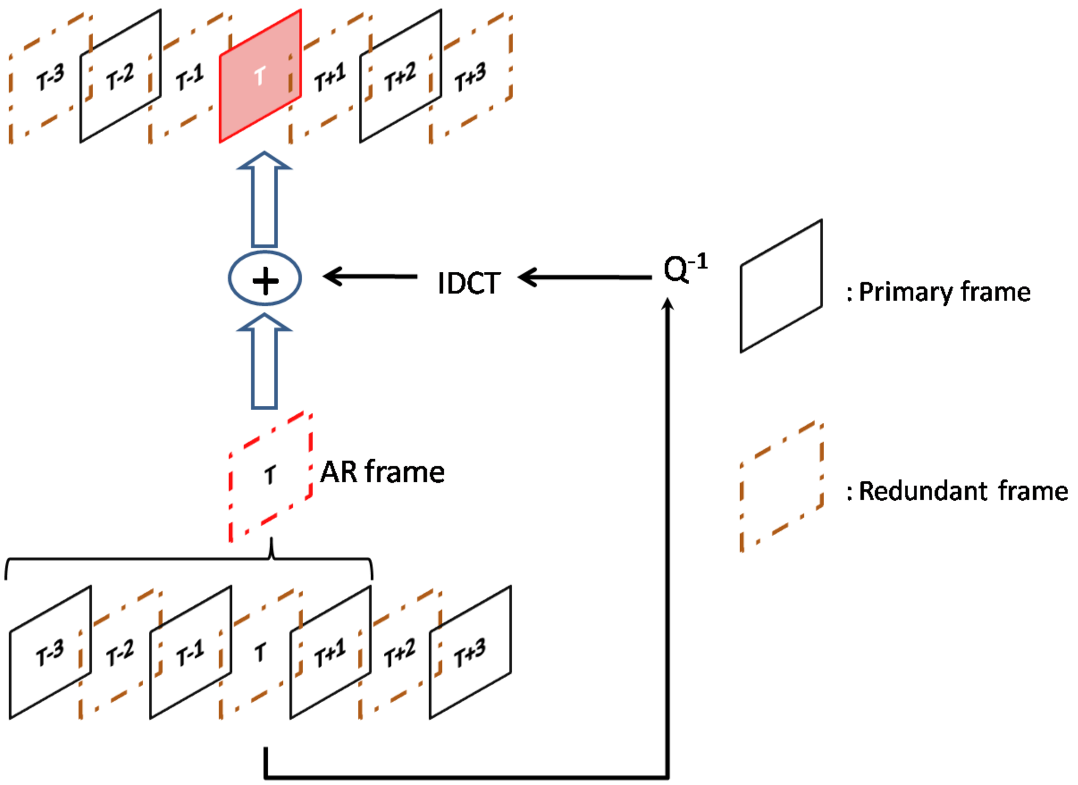
Loss cases	Description			Strategy
	Primary Frame T	Primary Frame T+1	Redundant Frame T	
No Primary Loss	No Loss	-	-	Decode directly
Single Primary Loss	Loss	No Loss	No Loss	AR frame + Residual
Primary and Redundancy Loss	Loss	No Loss	Loss	AR frame
Consecutive Primary Loss	Loss	Loss	-	Whole frame loss concealment

Table 3-1 Decoder Strategy under Different Packet Loss Circumstances

For the case of no primary loss, since the to-be-decoded primary frame T is received correctly, it is decoded directly. For the case of single primary loss, since the primary frame T is corrupted, and there is no further consecutive loss, the auto-regressive model is activated after decoding frame T+1. We obtain the interpolated AR frame of frame T by adopting auto-regressive model on frames T-3, T-1, and T+1, as illustrated in Figure 3-2(a), and improve the quality of AR frame with residual information i.e., the redundant frame of frame T, if it is available, as we illustrated in Figure 3-2(b).



(a) Produce AR Frame



(b) Improve AR Frame with Residual Information

Figure 3-2 Decoding Strategy for the Case of Simple Primary Loss

The residual information improvement is to extract residual by inverse residual coding which comprises inverse quantization and inverse DCT and then add it back to AR frame.

For the case of single primary and redundancy loss, it is similar to the case of single primary loss. The difference between two cases is AR-frame improving is not implemented after producing AR frame due to the lack of redundant frame as in Figure 3-2 (a).

For the case of consecutive primary loss, since there is consecutive loss on frame T+1, as in Figure 3-3, the self-feedback auto-regressive model cannot be performed.

Hence, other strategy for whole frame loss error concealment such as frame copy is employed.

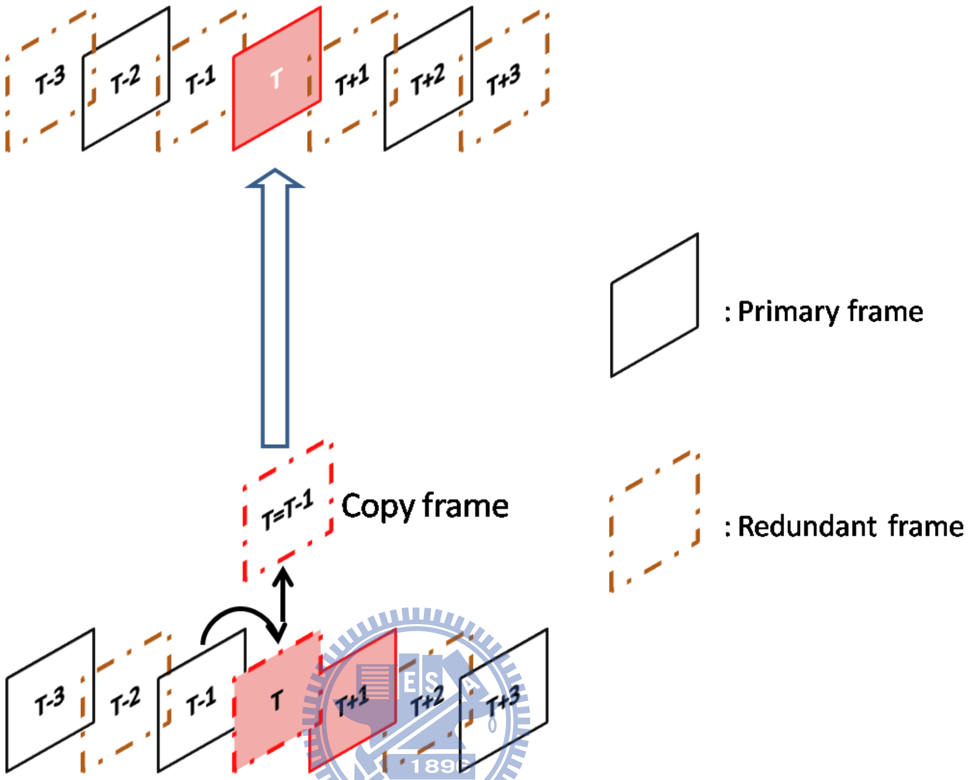


Figure 3-3 Decoding Strategy under Consecutive Loss

3.4 TAR Model

TAR model proposed by Zhang *et al* [6] is a compact version of STAR model. Due to its low complexity and good performance, temporal auto-regressive model (TAR model) is adopted to our proposal. The main difference between STAR model and TAR model is the support region selection.

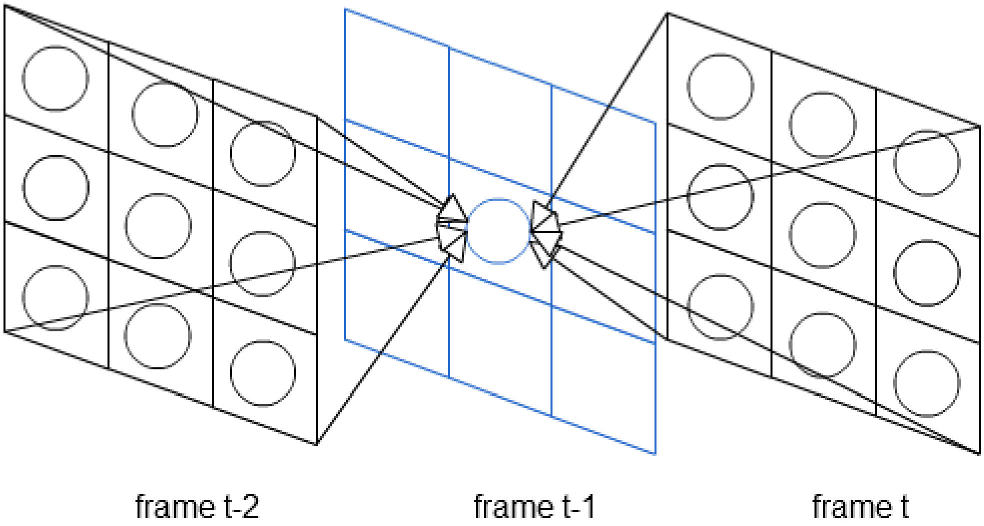


Figure 3-4 TAR Model Diagram

Figure 3-4 illustrates the TAR model diagram. According to the auto-regressive model, each pixel in the to-be-interpolated frame t-1 is modeled as a linear combination of only temporal neighboring pixels. In addition, all pixels in the same training window share the same weighting coefficients.

The pixel value of the to-be-interpolated frame t-1 is $\hat{R}_{t-1}$, which is defined as in (7).

$$\hat{R}_{t-1}(k,l) = \sum_{-L\leq(u,v)\leq L}\sum R_{t-2}(k+u,l+v)\times W_p(u,v) + \sum_{-L\leq(u,v)\leq L}\sum R_t(k+u,l+v)\times W_f(u,v) \tag{7}$$

The index t-1 represents the frame t-1. W_p, W_f stand for weighting coefficients in previous and following temporal neighborhood. The (k,l) is the pixel location in the training window, and (u,v) is the looping index for each temporal support region. In Figure 3-4, we assume that support order is 1 i.e., the support region is 3x3. Hence, the length of weighting vector will be 18.

The self-feedback weight training algorithm in TAR model is similar to the one in

STAR model. Due to the absence of spatial neighborhood, the formula (3) is simplified to adapt our modification as shown in (8).

$$\hat{R}_t(k,l) = \sum_{-L \le (u,v) \le L} \sum \hat{R}_{t-1}(k+u, l+v) \times W_p(u,v) + \sum_{-L \le (u,v) \le L} \sum \hat{R}_{t+1}(k+u, l+v) \times W_f(u,v) \quad \textbf{(8)}$$

Chapter 4 Experimental Results

4.1 Testing Environment and Model Parameters

To evaluate the performance of the proposed method (TAR-R), three more methods are used for comparison: *Scalable MDC with side information* (MCI-R) in [2], *Redundant intra coded frame* (HQP), and *MDC with STAR side information* (STAR-R). The STAR-R method is the variation of the proposed method. It substitutes TAR to STAR in the proposed model. For simulation purposes JM reference software version 15.1 [9] was used. Four test sequences: foreman, news, mobile and coastguard with QCIC (176x144) resolution are used for performance evaluation. The intra period is 10 frames for each descriptor; primary QP is 20 and redundancy rate is allocated up to half of primary rate. The four MDC methods used in the experiments are examined for their packet-loss performance, side-decoder performance, as well as the error propagation effect.

In TAR-R and STAR-R, we adopted MCI with block size 8x8 and search window size 4x4 in quarter pixel accuracy to be the initial value of AR model. Full search was applied in bi-direction motion estimation here. The same MCI frames are used as the interpolated frames in MCI-R method.

Parameters for regression models in TAR-R and STAR-R are listed below:

- Regression window size Wx, Wy:

  QCIF: 16x16

- Maximum iteration times:

  2 for both STAR and TAR model

- Support order

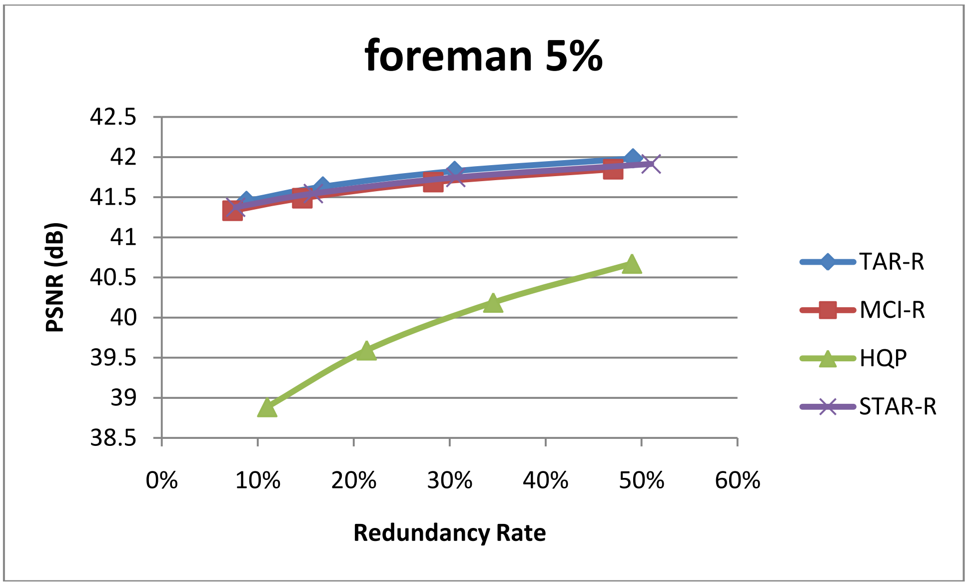
  1 in STAR and TAR model

4.2 Packet Loss Performance

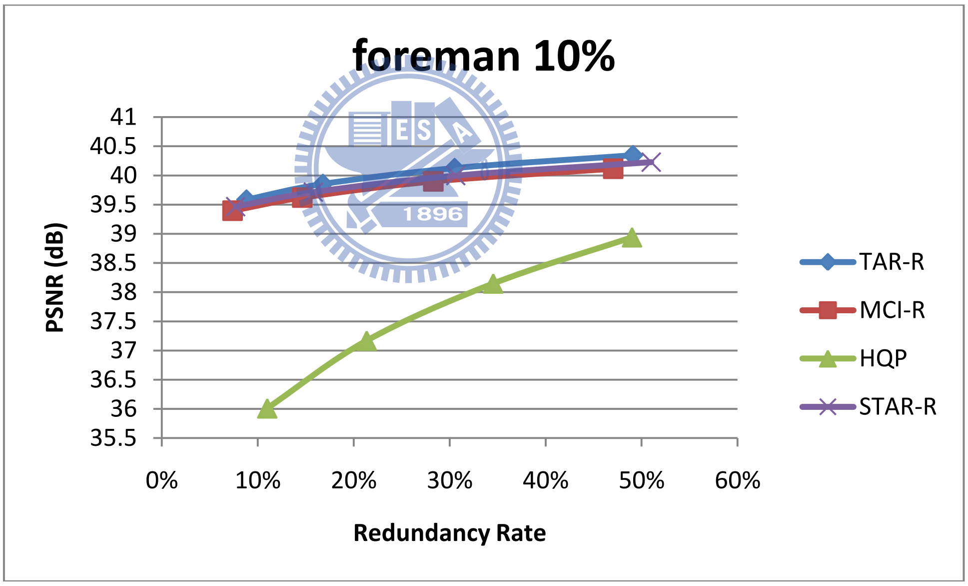
The four MDC methods were examined in a packet-loss scenario where various packet-loss rates, 5%, 10%, and 20% are adopted. We use one packet for each frame of each descriptor. Packets are lost randomly in identical independent distributed Bernoulli process.

4.2.1 General Sequences

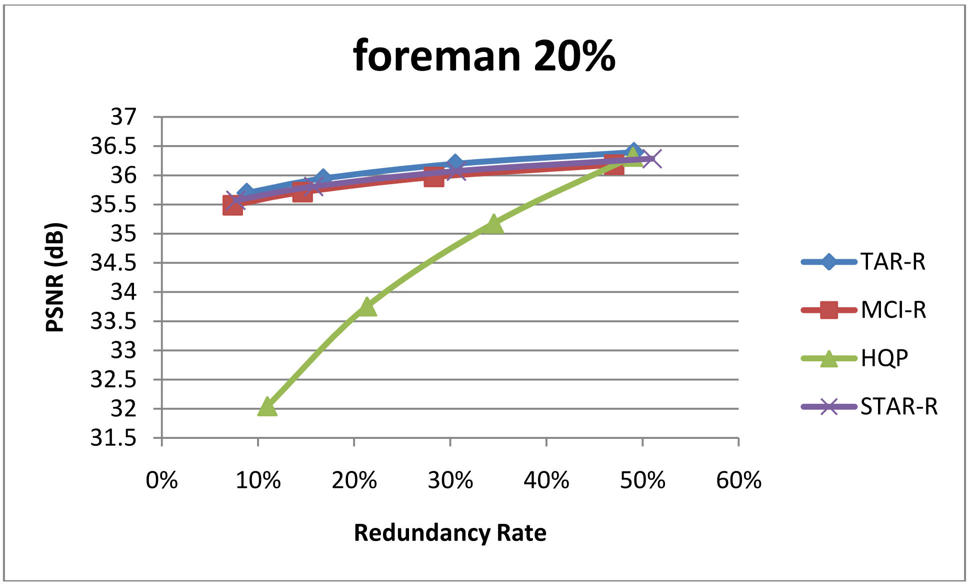
Foreman is a video sequence with median-motion content. We first examined the performance of the four MDC methods on it under packet-loss scenario and the result is shown in Figure 4-1, where the rate-distortion curves under different packet loss rates are presented. The result shows that MDC with residual frames such as MCI-R, TAR-R and STAR-R have better coding efficiency. When the packet loss rate becomes higher, the PSNR gap between HQP and other methods gets smaller because HQP suffers no error propagation, while others are not error propagation-free. As shown in Figure 4-1 (c), under 20% packet loss rate, HQP performs as well as others when the redundancy rate is almost half of primary rate, which is costly.



(a) Foreman sequence at 5% Packet Loss Rate (QCIF)-Primary rate=1185.51kbps



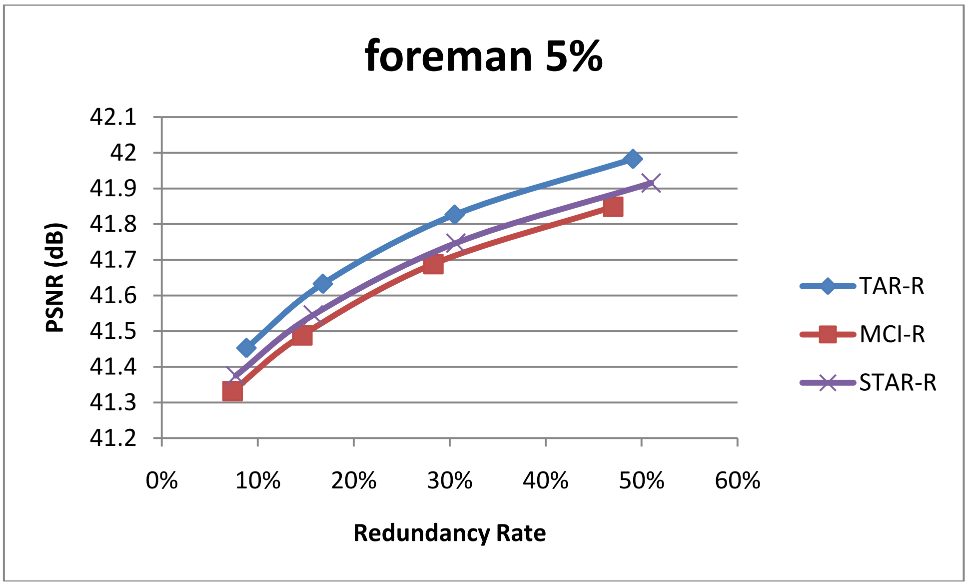
(b) Foreman sequence at 10% Packet Loss Rate (QCIF) -Primary rate=1185.51kbps



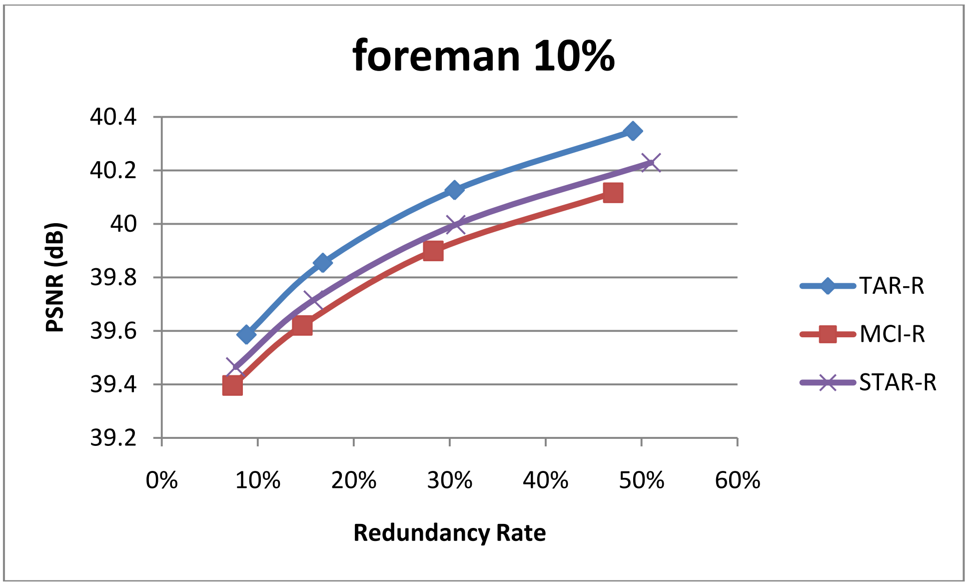
(c) Foreman sequence at 20% Packet Loss Rate (QCIF) -Primary rate=1185.51kbps

Figure 4-1 Rate-Distortion Performance Comparison in Packet-Loss Environment (Foreman / Full)

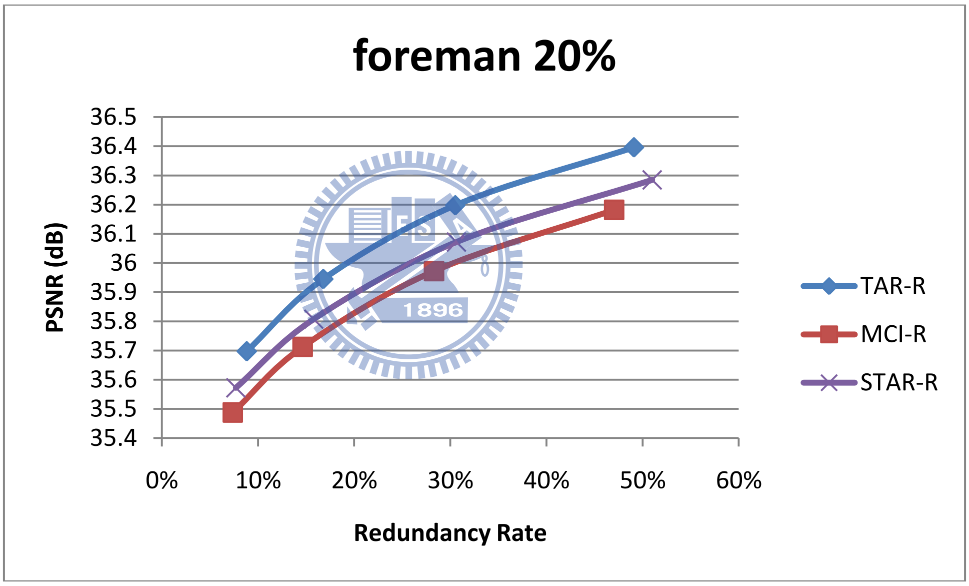
To compare methods beside HQP, we remove the curve of HQP in Figure 4-2 to make the curves of other methods more clear. It is observed that our proposed method TAR-R has the best performance among others due to the better interpolated frame.



(a) Foreman sequence at 5% Packet Loss Rate (QCIF) -Primary rate=1185.51kbps



(b) Foreman sequence at 10% Packet Loss Rate (QCIF) -Primary rate=1185.51kbps



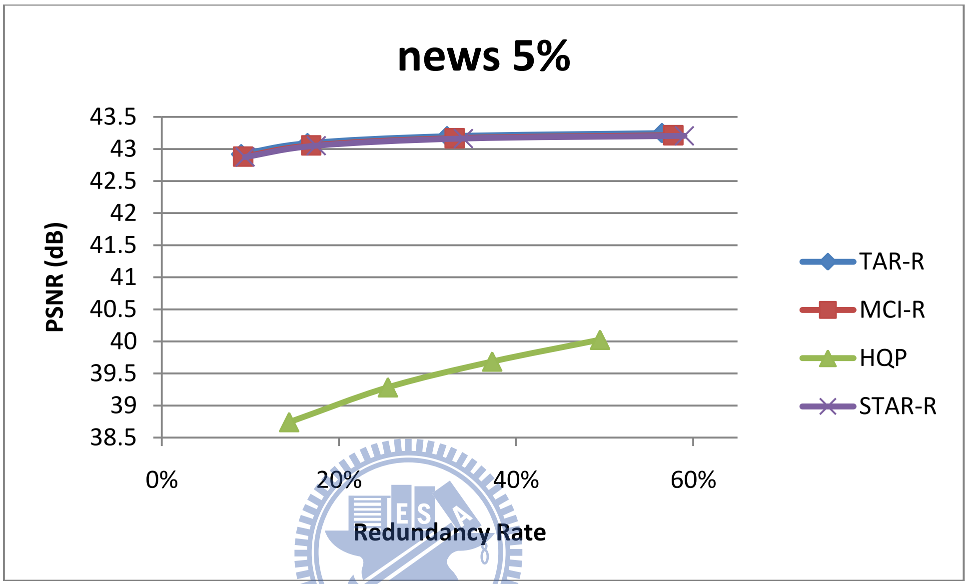
(c) Foreman sequence at 20% Packet Loss Rate (QCIF) -Primary rate=1185.51kbps

Figure 4-2 Rate-Distortion Performance Comparison in Packet-Loss Environment (Foreman / Partial)

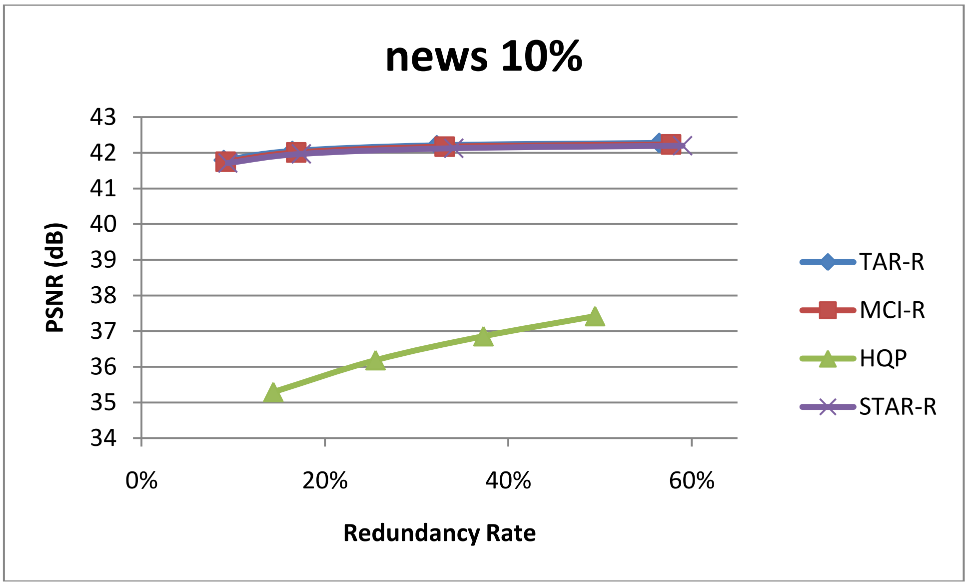
4.2.2 Static Sequences

News is a video sequence with low-motion content. We use News sequence to see how well each MDC method performs with static and low-motion content and the result is depicted in Figure 4-3. From the figure, it can be seen that the PSNR gap between

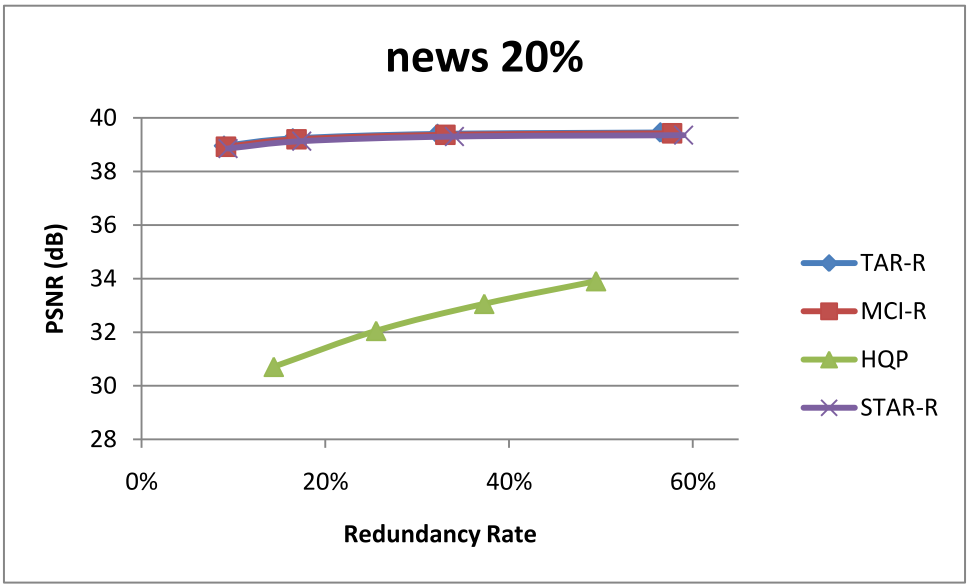
HQP and others is still obvious. Owing to low-motion, AR frames produced in TAR-R and STAR-R methods and the interpolated frames produced in MCI-R method are all in good quality. Thus, the residual coding bit rate are reduced in these methods, resulting in large PSNR gap between these methods and HQP.



(a) News sequence at 5% Packet Loss Rate (QCIF)-Primary rate=709.94 kbps



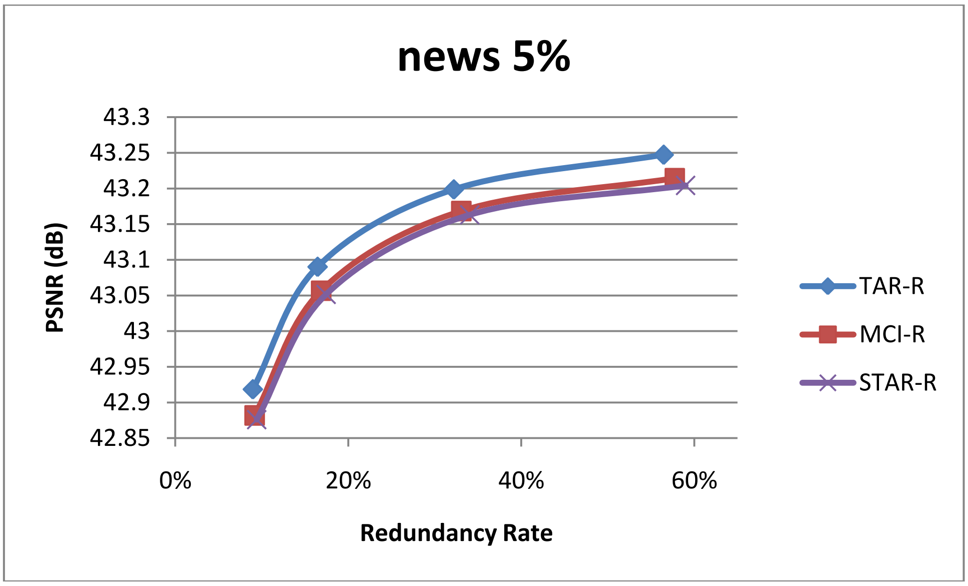
(b) News sequence at 10% Packet Loss Rate (QCIF) -Primary rate=709.94 kbps



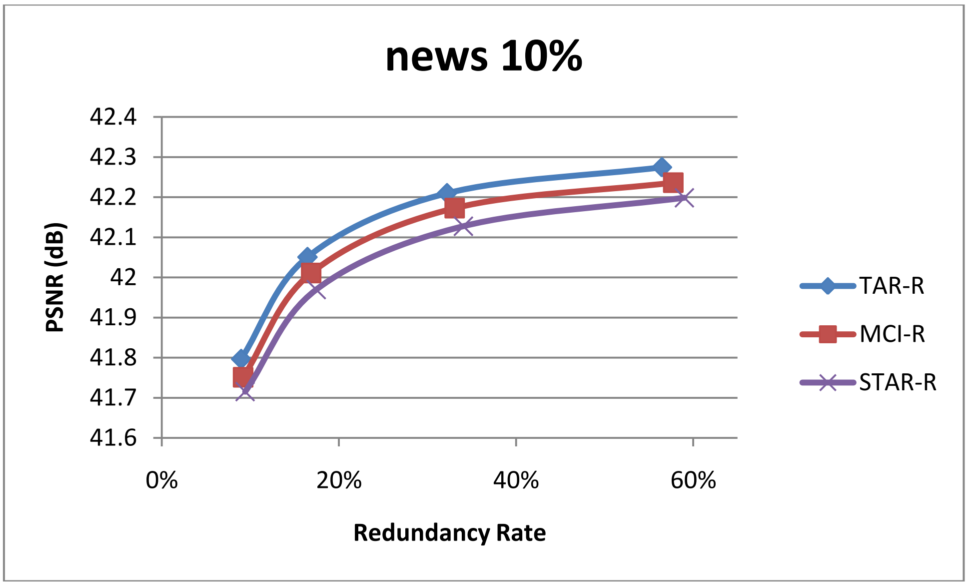
(c) News sequence at 5% Packet Loss Rate (QCIF) -Primary rate=709.94 kbps

Figure 4-3 Rate-Distortion Performance Comparison in Packet-Loss Environment (News / Full)

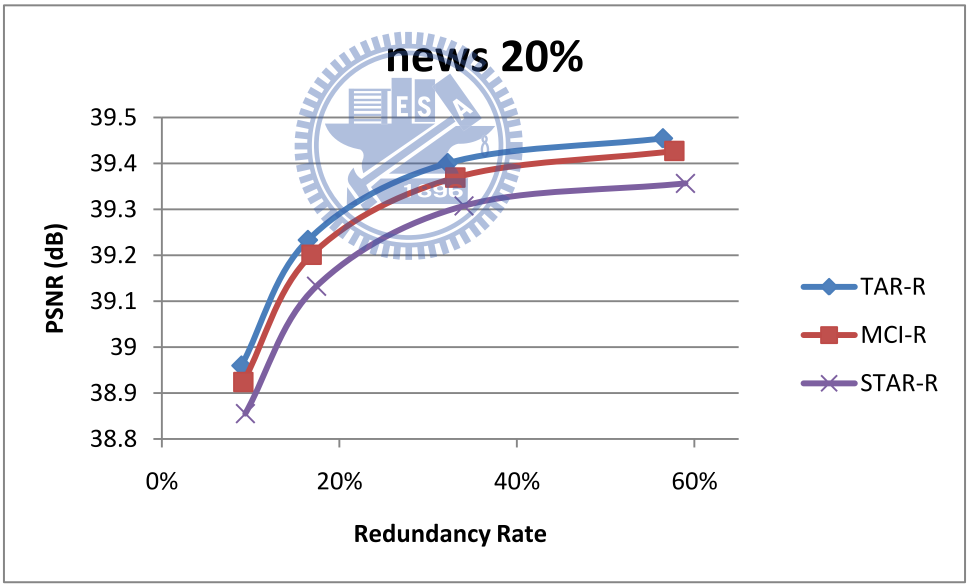
By removing HQP curve, the detailed comparison between TAR-R, STAR-R and MCI-R is illustrated in Figure 4-4. The overall results of the three methods are very close. The proposed TAR-R performs better than others with only little PSNR gain.



(a) News sequence at 5% Packet Loss Rate (QCIF) -Primary rate=709.94 kbps



(b) News sequence at 10% Packet Loss Rate (QCIF) -Primary rate=709.94 kbps

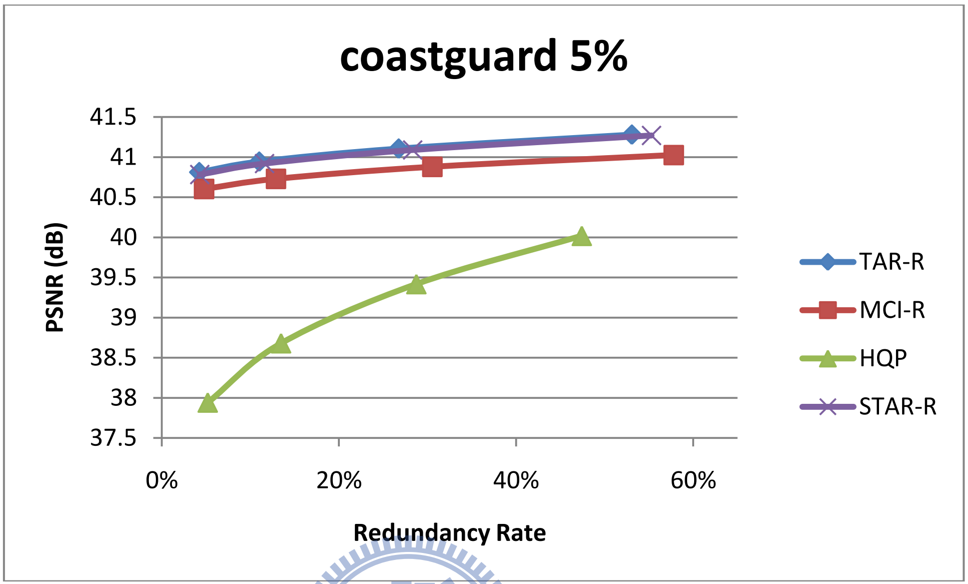


(c) News sequence at 5% Packet Loss Rate (QCIF) -Primary rate=709.94 kbps

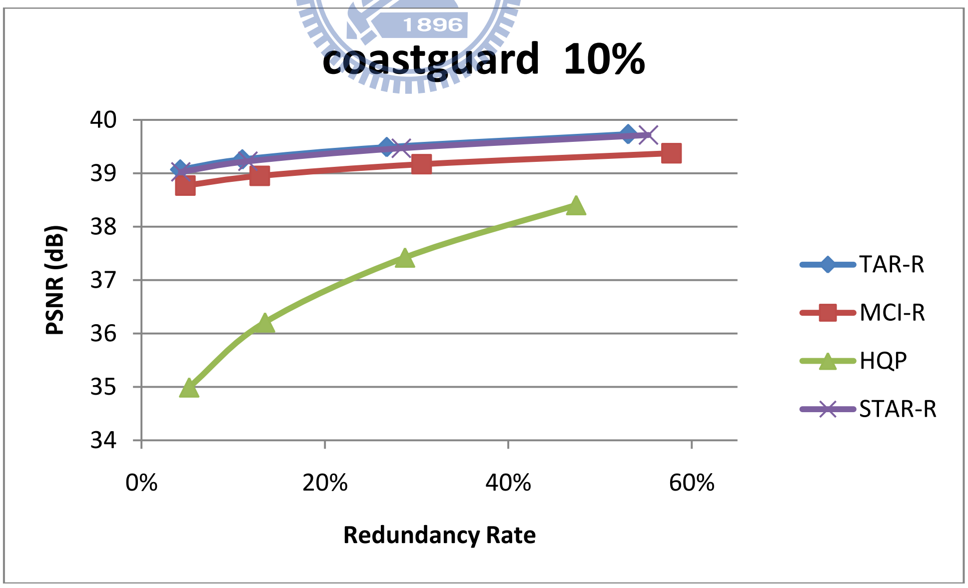
Figure 4-4 Rate-Distortion Performance Comparison in Packet-Loss Environment (News / Partial)

4.2.3 Dynamic Sequences

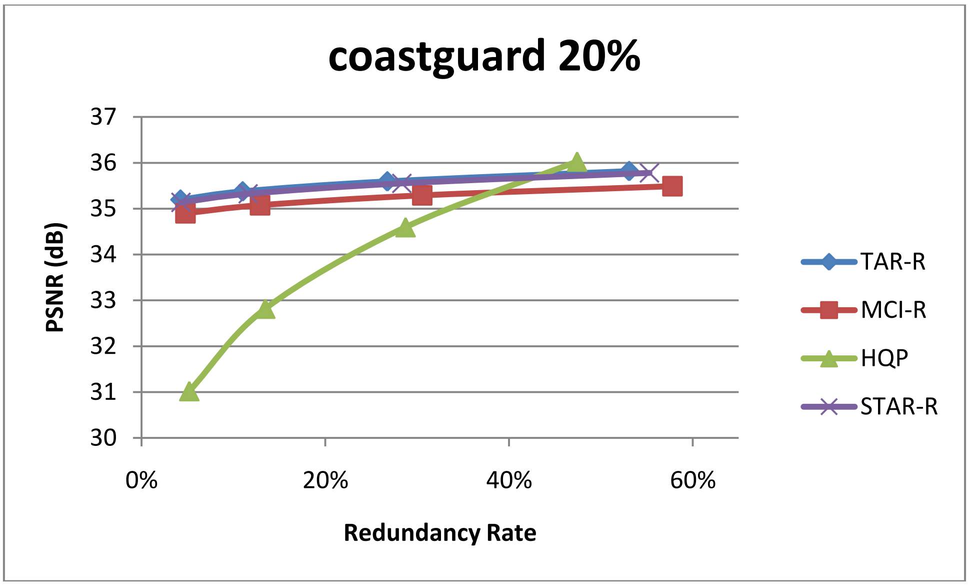
We use coastguard sequence to assess the performance of the four MDC methods when the coding sequence is more dynamic with higher motion. As in Figure 4-5, the result is close to that of using foreman sequence. Compared to static sequences such as news we tested before, sequences with higher and motion, the PSNR gap between HQP and others is less obvious as the curves crossed in high loss rate.



(a) Coastguard sequence at 5% Packet Loss Rate (QCIF)-Primary Rate=2165.98 kbps



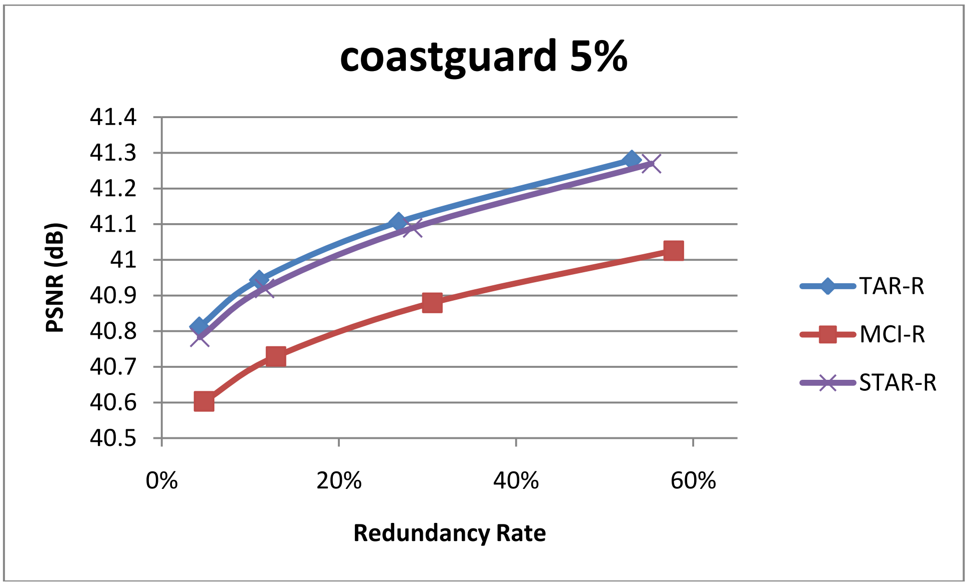
(b) Coastguard sequence at 10% Packet Loss Rate (QCIF) -Primary Rate=2165.98 kbps



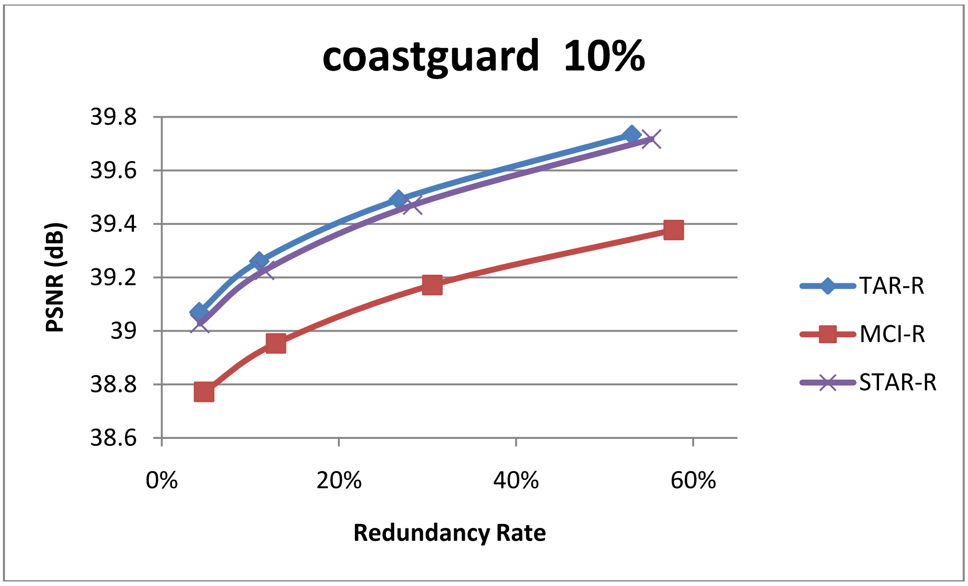
(c) Coastguard sequence at 5% Packet Loss Rate (QCIF) -Primary Rate=2165.98 kbps

Figure 4-5 Rate-Distortion Performance Comparison in Packet-Loss Environment (Coastguard / Full)

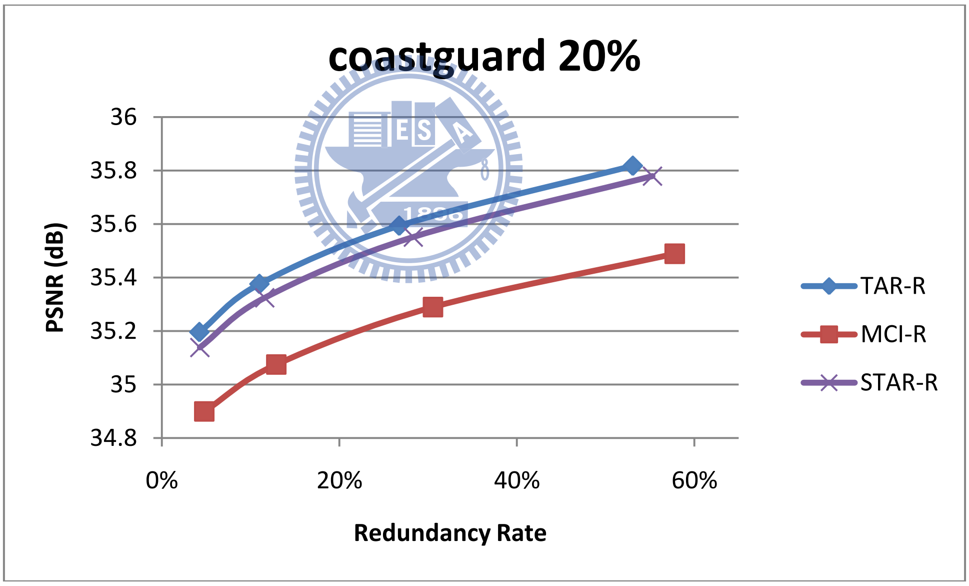
By removing HQP curve, Figure 4-6 presents detailed comparison between TAR-R, STAR-R and MCI-R for coastguard sequence. It is observed that both TAR-R and STAR-R perform much better than MCI-R and that the performance difference between TAR-R and STAR-R is not obvious. Since TAR-R reduces complexity and produces even better quality than STAR-R, the result demonstrates the superiority of TAR-R. .



(a) Coastguard sequence at 5% Packet Loss Rate (QCIF) -Primary Rate=2165.98 kbps



(b) Coastguard sequence at 10% Packet Loss Rate (QCIF) -Primary Rate=2165.98 kbps



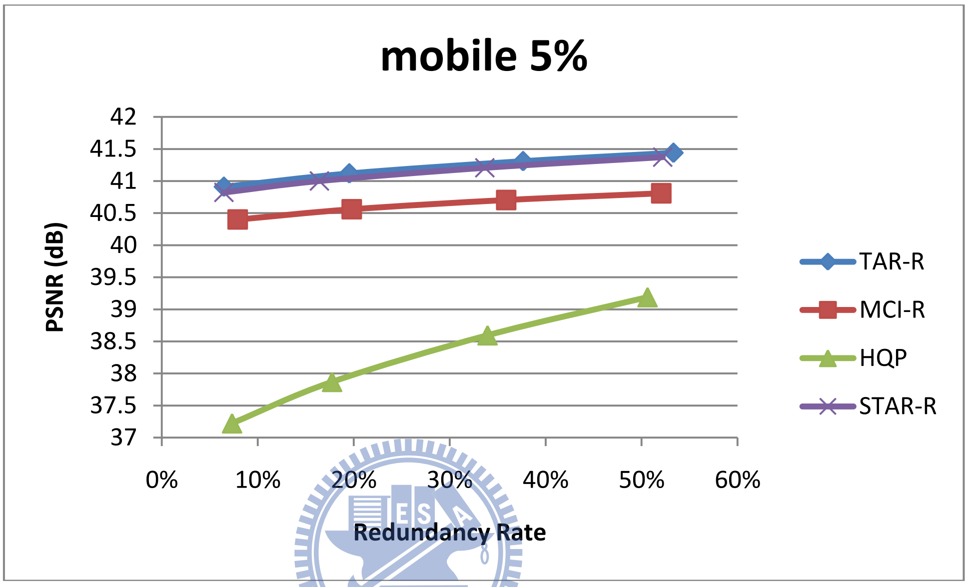
(c) Coastguard sequence at 5% Packet Loss Rate (QCIF) -Primary Rate=2165.98 kbps

Figure 4-6 Rate-Distortion Performance Comparison in Packet-Loss Environment (Coastguard / Partial)

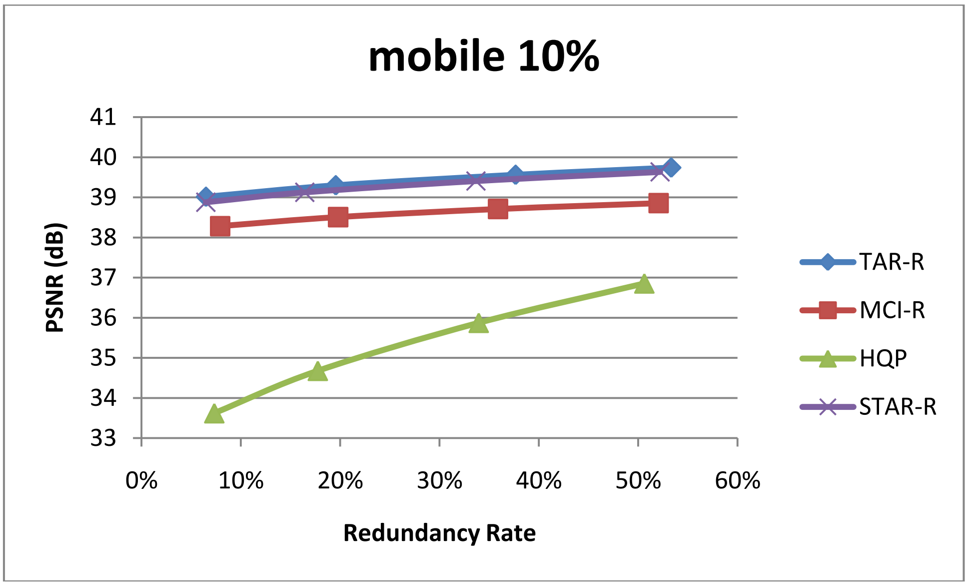
4.2.4 Complicated Sequences

Beside static and dynamic sequences, we further observed the behavior of the four MDC methods when the coding sequences are more complicated. We use mobile sequence as a benchmark because there are zooming and more complicated motion in the sequence.

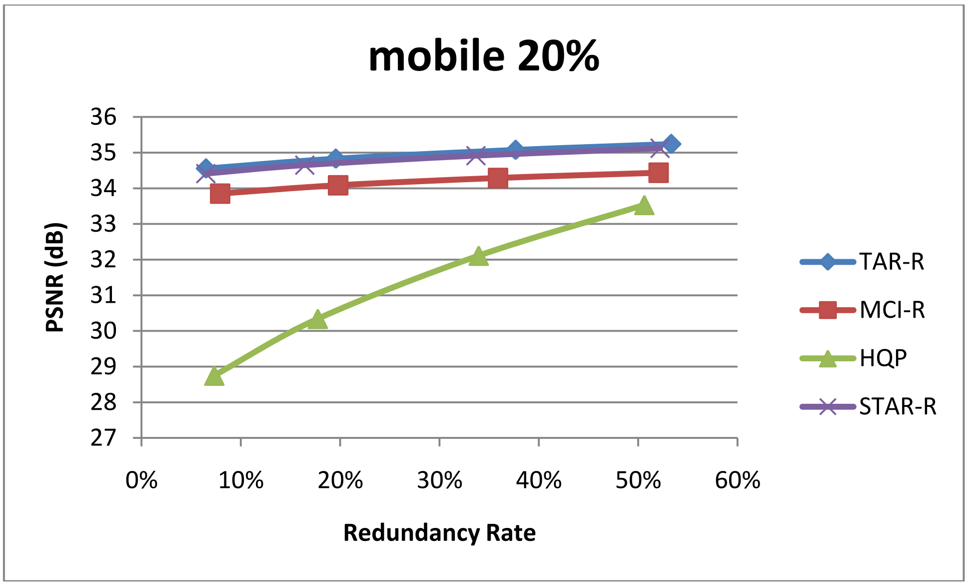
As in Figure 4-7 HQP performs worse than other methods, and it is consistent to the result of other test sequences we examined before.



(a) Mobile sequence at 5% Packet Loss Rate (QCIF)-Primary Rate=2891.58 kbps



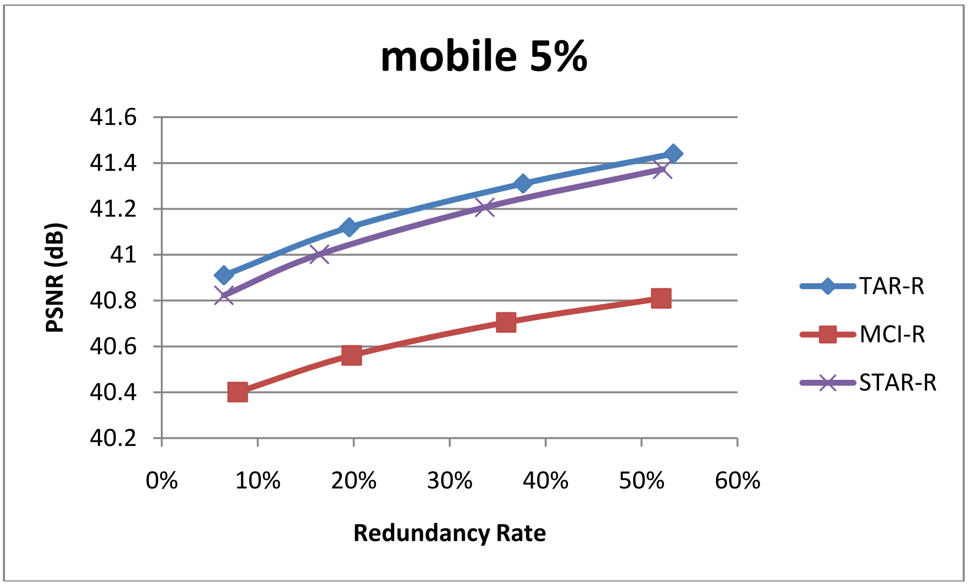
(b) Mobile sequence at 10% Packet Loss Rate (QCIF) -Primary Rate=2891.58 kbps



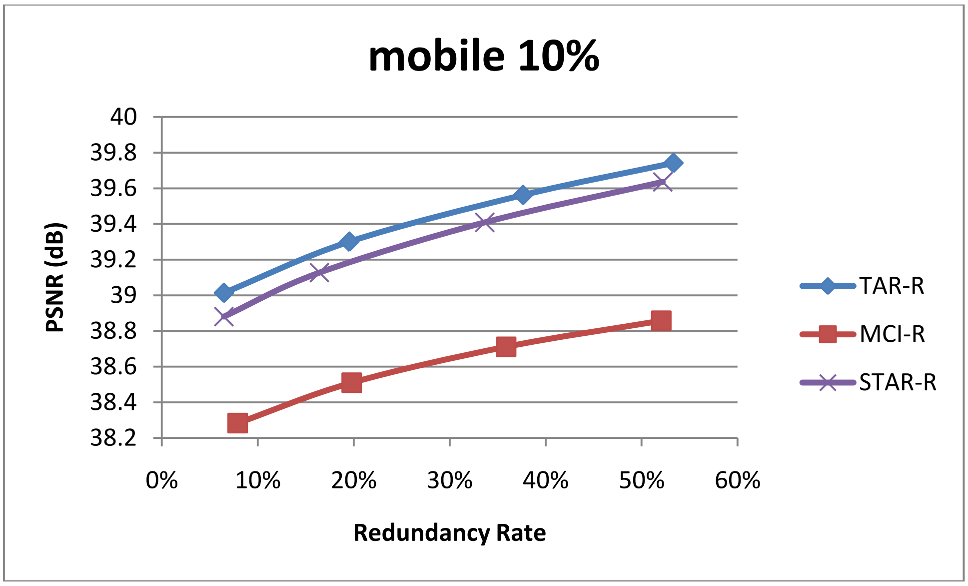
(c) Mobile sequence at 20% Packet Loss Rate (QCIF) -Primary Rate=2891.58 kbps

Figure 4-7 Rate-Distortion Performance Comparison in Packet-Loss Environment (Mobile / Full)

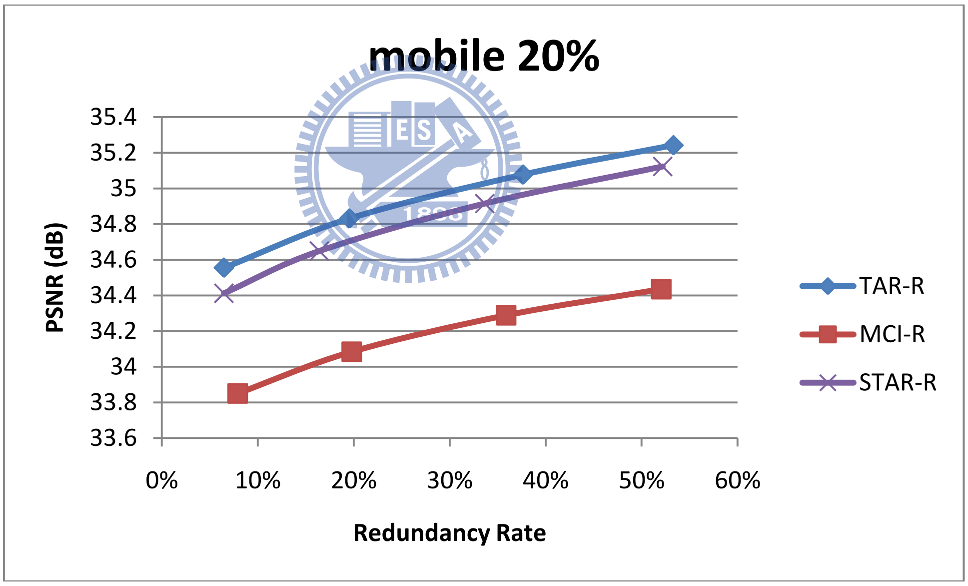
By removing HQP curve, we compared TAR-R MCI-R and STAR-R in Figure 4-8. The reason why TAR-R and STAR-R beat MCI-R more than the cases with other test sequences is because they take advantage of using auto-regressive model which improve the quality of interpolated frames especially when it comes to zooming or panning.



(a) Mobile sequence at 5% Packet Loss Rate (QCIF) -Primary Rate=2891.58 kbps



(b) Mobile sequence at 10% Packet Loss Rate (QCIF) -Primary Rate=2891.58 kbps

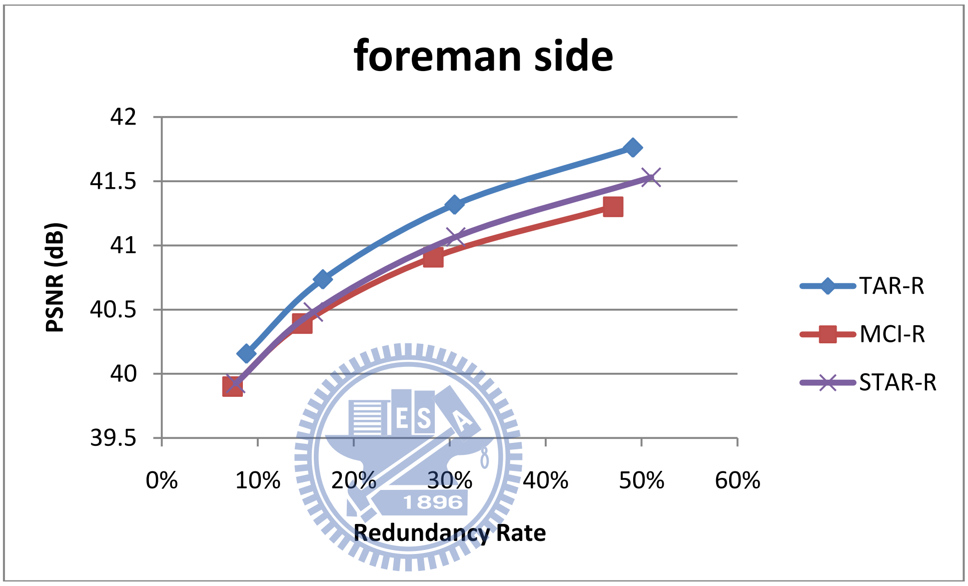


(c) Mobile sequence at 20% Packet Loss Rate (QCIF)-Primary Rate=2891.58 kbps

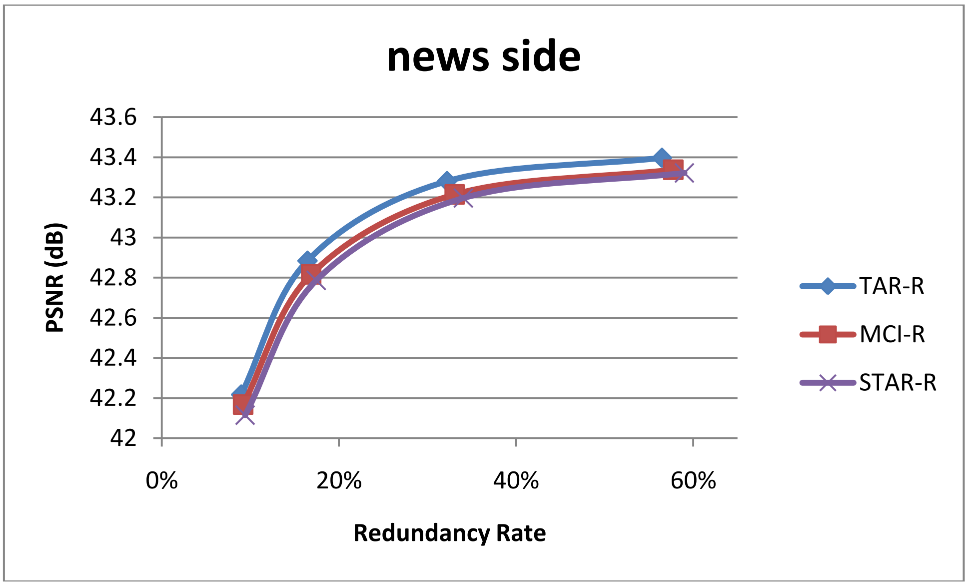
Figure 4-8 Rate-Distortion Performance Comparison in Packet-Loss Environment (Mobile / Partial)

4.3 Side-Decoder Performance

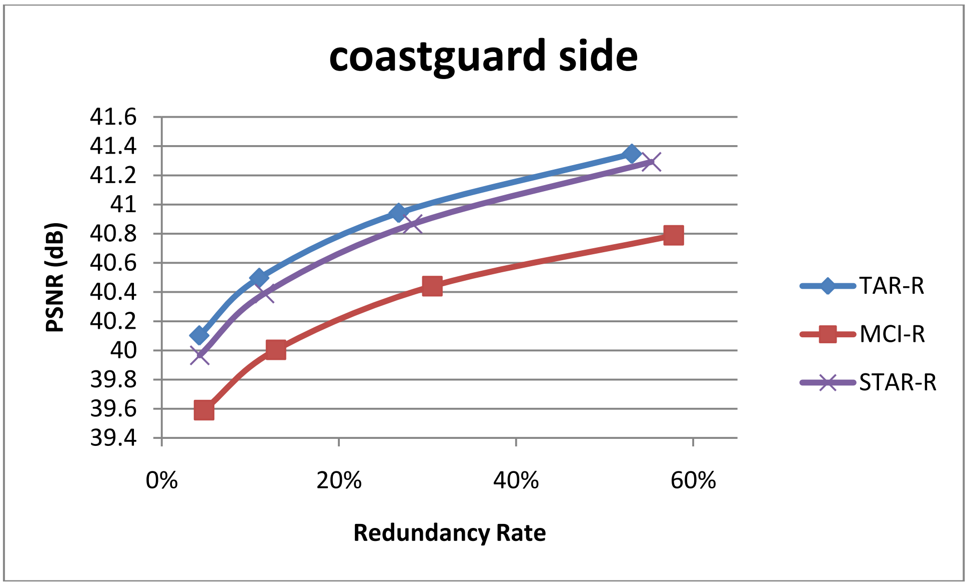
In this section we examined the four MDC methods with side-decoder performance. The results of two side decoders are averaged and shown in Figure 4-9. For the case of foreman, TAR-R is overall the best. For the case of coastguard and mobile sequences, TAR-R and STAR-R perform better than MCI-R while for the case of news sequence, three curves didn't separate much since the concealed frames by motion compensated interpolation was already in good quality such that there is not much for auto-regressive model to improve in TAR-R and STAR-R.



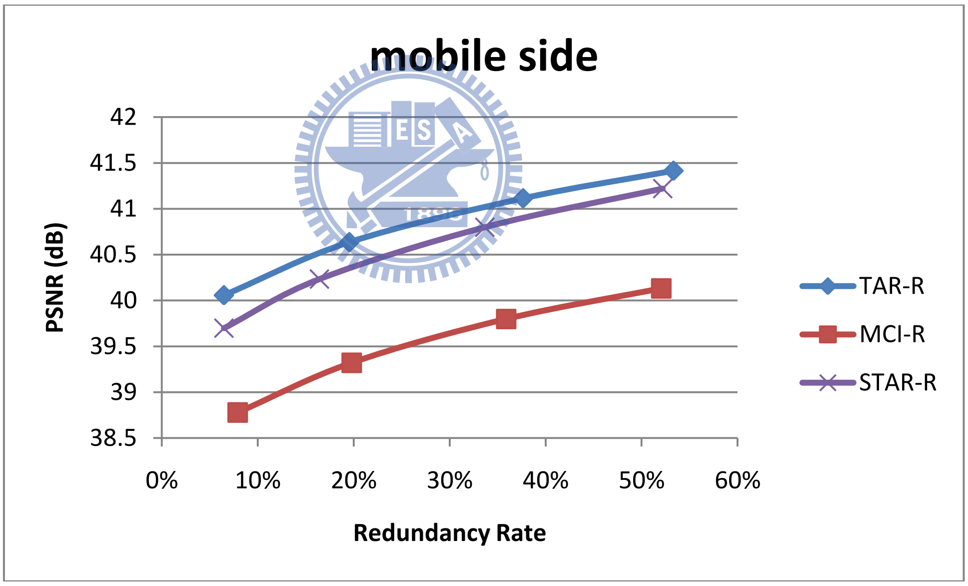
(a) Foreman sequence at side-decoder (QCIF) -Primary rate=1185.51kbps



(b) News sequence at side-decoder (QCIF) -Primary rate=709.94 kbps



(c) Coastguard sequence at side-decoder (QCIF) -Primary Rate=2165.98 kbps



(d) Mobile sequence at side-decoder (QCIF) -Primary Rate=2891.58 kbps

Figure 4-9 Side Performance Comparison

4.4 Frame-by-frame distortion performance observation

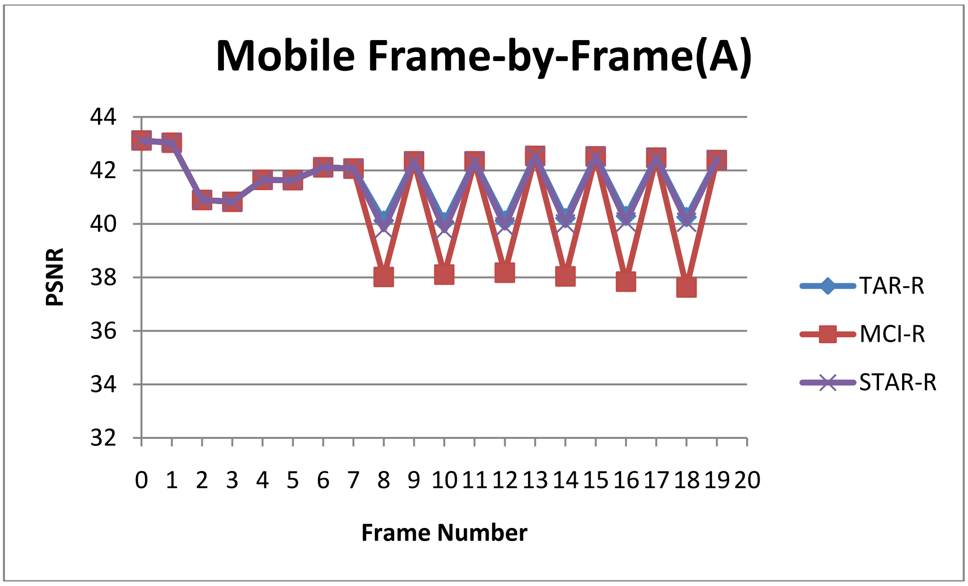
We set up a loss scenario and check the frame-by-frame distortion performance to observe the error propagation effect of different comparison methods. For the simulation, we use mobile sequence as our test sequence. According to the PSNR gap between HQP and other methods, we ignore HQP here.

The first group of pictures we set up a case of single primary loss where primary frame is lost at frame 8 as shown in Figure 4-10 (a).

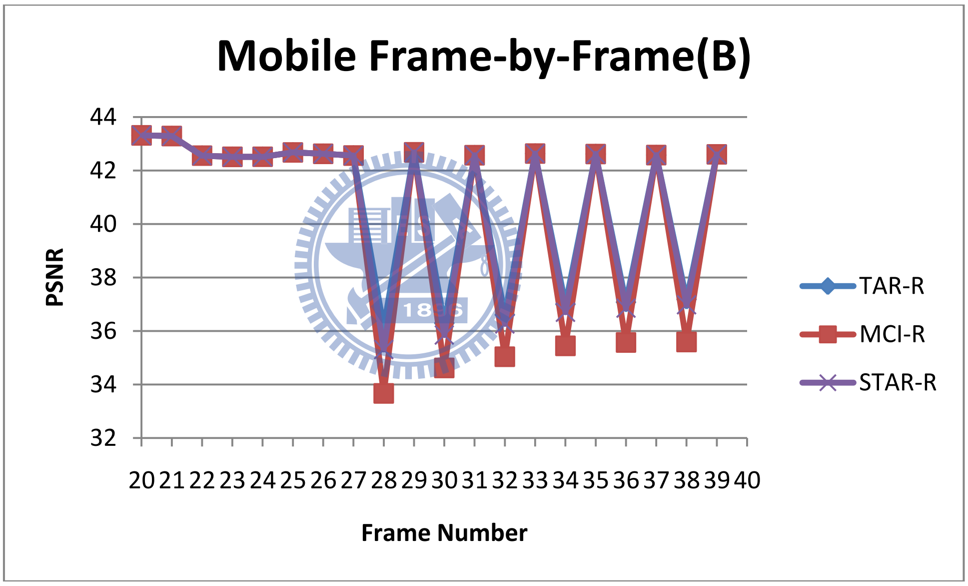
The second group of pictures we set up a case of primary and redundancy loss where primary frame and redundant frame are lost at frame 28 as in Figure 4-10 (b). Frame 28 is then concealed to interpolated frame, i.e., AR frame in TAR-R and STAR-R, and motion compensated interpolation frame in MCI-R.

Figure 4-10 (c) shows the third group of pictures where the primary frames are lost at frame 48 and frame 51. Both frame 48 and frame 51 would be concealed as a single primary loss case. For TAR-R and STAR-R, AR frame 51 will be generated by auto-regressive model with frame 48, frame 50 and frame 52, which are corrupted and error propagated frames and then improved by adding the residual. For MCI-R, interpolated frame will be generated by frame 50 and frame 52, which are also error propagated frames. Hence, we set up this scenario to observe the performance when the three methods conceal a frame by corrupted frames.

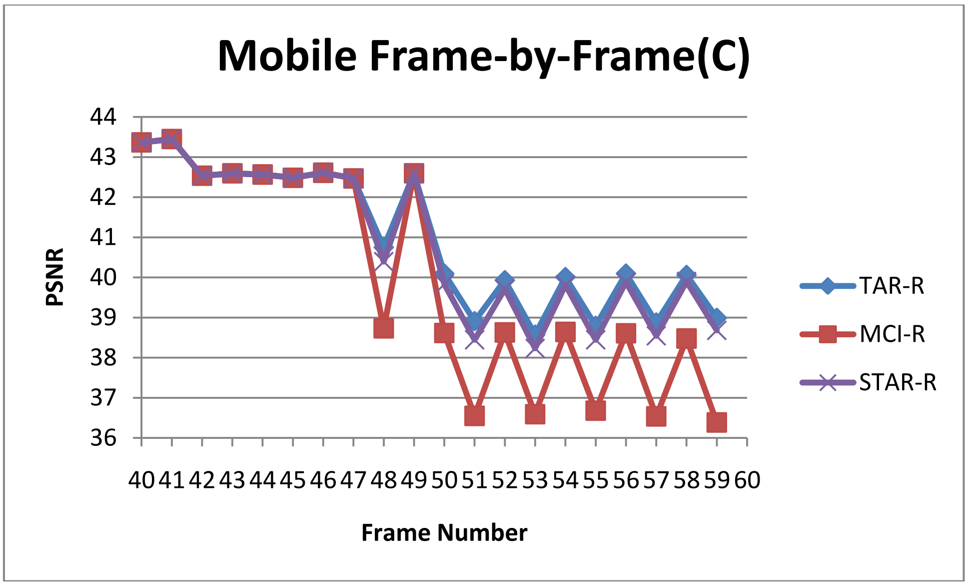
With these three cases, our proposed method TAR-R has overall the best performance.



(a) Single Primary Loss



(b) Primary and Redundancy Loss



(c) Frame concealment by corrupted frame

Figure 4-10 Mobile Sequence Frame-by-Frame Performance (QCIF)

Chapter 5 Conclusion

In this paper we propose an auto-regressive model enhanced multiple description coding. The model produces two symmetric descriptions in primary and redundant frames interleaved manner by each frame. Simplifying the STAR model, we adopted TAR model into our MDC architecture which carries redundant frames predicted by temporal AR frames. We compared our proposal with three other similar methods: Redundant intra coded frame (HQP), MDC with STAR side information (STAR-R), and Scalable MDC with side information (MCI-R) by experiments. The experimental results show that MDC with residual frames is more coding efficient than MDC with redundant intra frames. Furthermore, with the aid of auto-regressive model, our proposed method (TAR-R) and STAR-R gains more coding efficiency in residual frame coding than MCI-R did. Inspecting the case of packet loss performance, side-decoder performance, and error propagation resistance, we showed our proposal is the most promising architecture among these methods.

REFERENCE

[1] V. K. Goyal, “Multiple description coding: Compression meets the network,” IEEE Signal Process. Mag., vol. 18 no. 5 pp. 74-93, Sep. 2001

[2] S. Adedoyin, W. A. C. Fernando, H. A. Karim, C. T. E. R. Hewage, A. M. Kondoz, “Scalable Multiple Description Coding with Side Information Using Motion Interpolation”, Consumer Electronics, IEEE Transactions on, November 2008, VOL. 54, NO. 4.

[3] J. Apostolopoulos, “Reliable video communication over lossy packet networks using multiple state encoding and path diversity,” Proceedings of Visual Communications: Image Processing, pp. 392-409, Jan. 2001.

[4] N. Franchi, M. Fumagalli, R. Lancici and S. Tubaro, “Multiple description video coding for scalable and robust transmission over IP,” Circuits and Systems for Video Technology, IEEE Transactions on, March 2005, VOL.15, NO.3

[5] A. R. Reibman, H. Jafarkhani, M. T. Orchard and Y. Wang, “Multiple description video using rate-distortion splitting,” Proceedings of IEEE International Conference on Image Processing, Oct. 2001, pp. 978-981

[6] Yongbing Zhang, Debin Zhao, Xiangyang Ji, Ronggang Wang, and Wen Gao, “A Spatio-Temporal Auto Regressive Model for Frame Rate Upconversion”, Circuits and Systems for Video Technology, IEEE Transactions on, September 2009, VOL.

19, NO. 9.

[7] V.A. Vaishampayan, “Design of Multiple Description Scalar Quantizers,” IEEE Transaction on Information Theory, vol. 39, 1993.

[8] Tammam Tillo, Marco Grangetto, Gabriella Olmo, “Redundant Slice Optimal Allocation for H.264 Multiple Description Coding” Circuits and Systems for Video Technology, IEEE Transactions on, January 2008, VOL.18, NO.1

[9] H.264/AVC Ref. Software – JM, http://iphome.hhi.de/suehring/tml/.