การปรับปรุงเทคนิคการปกปิดความผิดพลาดสำหรับการเข้ารหัสวีดิทัศน์ H.264 แบบปรับขนาดได้

นายซิมอน จูด กัว ลาม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

IMPROVEMENT OF ERROR CONCEALMENT TECHNIQUE FOR H.264 SCALABLE VIDEO CODING

Mr. Simon Jude Que Lam

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2009

Thesis Title	IMPROVEMENT OF ERROR CONCEALMENT TECHNIQUE FOR H.264 SCALABLE VIDEO CODING
By	Mr. Simon Jude Que Lam
Field of Study	Electrical Engineering
Thesis Advisor	Assistant Professor Supavadee Aramvith, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Engineering

(Associate Professor Boonsom Lerdhirunwong, Dr.Ing)

THESIS COMMITTEE

.... Chairman

(Professor Prasit Prapingmongkolkarn, Ph.D.)

.... Thesis Advisor

(Assistant Professor Supavadee Aramvith, Ph.D.)

.... Examiner

(Assistant Professor Charnchai Pluempitiwiriyawej, Ph.D.)

.... External Examiner

(Assistant Professor Thanarat Chalidabhongse, Ph.D.)

ซิมอน จูด กัว ลาย: การปรับปรุงเทคนิคการปกปิดความผิดพลาดสำหรับการเข้ารหัสวีดิทัศน์ H.264 แบบปรับขนาดได้ (IMPROVEMENT OF ERROR CONCEALMENT TECHNIQUE FOR H.264 SCALABLE VIDEO CODING) อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ .ดร.สุภาวดี อร่ามวิทย์, 60 หน้า.

ในปัจจุบันเทคโนโลยีการเข้ารหัสวีดิทัศน์แบบปรับขนาดได้นำมาซึ่งคำตอบของการส่งคอนเทนต์สื่อประสมไปยังอุปกรณ์ปลายทางบนโครงข่ายหลากหลายรูปแบบ ในวิทยานิพนธ์นี้ประเด็นหลักอยู่ที่งานประยุกต์การปรับขนาดได้เชิงพื้นที่ ซึ่งสนับสนุนการส่งคอนเทนต์สื่อประสมไปยังผู้ใช้ที่มีอุปกรณ์ที่สามารถแสดงผลวีดิทัศน์ความละเอียดต่ำไปจนถึงวีดิทัศน์ความละเอียดสูง อย่างไรก็ตามการส่งข้อมูลผ่านโครงข่ายยังคงใช้การส่งแบบแพ็คเก็ตซึ่งไม่สามารถรับรองคุณภาพของบริการได้ ดังนั้นการสูญเสียของแพ็คเก็ตจะทำให้กระแสบิตวีดิทัศน์ถูกทำลาย ในกรณีของวีดิทัศน์ H.264 ซึ่งมีอัตราบีบอัดสูงนั้น การที่เฟรมวีดิทัศน์มีการสูญหายเพียงหนึ่งเฟรมอาจส่งผลกระทบต่อคุณภาพการมองเห็นอย่างมากต่อเฟรมที่ถูกกระทบและเฟรมต่อไป

เทคนิคการปกปิดความผิดพลาดมีการประยุกต์ใช้ด้านตัวถอดรหัสเพื่อประมาณข้อมูลที่สูญหายไป งานวิจัยที่เกี่ยวข้องกับเทคนิคการปกปิดความผิดพลาดของวีดิทัศน์ H.264 ได้ถูกนำเสนออย่างไรก็ดีประสิทธิภาพในบางกรณียังไม่เพียงพอ ในวิทยานิพนธ์นี้ได้นำเสนอการปรับปรุงกระบวนการการปกปิดความผิดพลาด โดยเสนอกระบวนวิธีการชดเชยการเคลื่อนที่โดยใช้ค่าเฉลี่ยแบบถ่วงน้ำหนักของเวคเตอร์การเคลื่อนที่ในมาโครบล็อกที่อยู่ในบริเวณเดียวกันจากเฟรมที่ถูกถอดรหัสก่อนหน้า อีกทั้งเฟรมที่ถูกสร้างขึ้นมาใหม่นี้จะถูกปรับปรุงอีกครั้งด้วยการกรองความถี่ผ่านช่วงกลางแบบลูกผสมเพื่อให้ประสิทธิภาพดีขึ้น

วิธีนำเสนอทำให้ค่าเฉลี่ยเพิ่มขึ้นถึง 0.8 เดซิเบล มากกว่าวิธีที่ดีที่สุดในปัจจุบัน ซึ่งถูกแสดงให้เห็นถึงคุณภาพที่ดีขึ้นในการประเมินค่าเชิงอัตวิสัย นอกจากนั้นยังแสดงให้เห็นว่าวิธีที่นำเสนอสามารถหยุดยั้งผลของการกระจายความผิดพลาดในเฟรมที่สูญหายได้อีกด้วย

ภาควิชา....วิศวกรรมไฟฟ้า........ ลายมือชื่อนิสิต..............................
สาขาวิชา......วิศวกรรมไฟฟ้า.... ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก..............
ปีการศึกษา 2552

5170665321: MAJOR ELECTRICAL ENGINEERING

KEYWORDS: ERROR CONCEALMENT/ SCALABLE VIDEO CODING / MOTION ESTIMATION / H.264/SVC

SIMON JUDE QUE LAM: IMPROVEMENT OF ERROR CONCEALMENT TECHNIQUE FOR H.264 SCALABLE VIDEO CODING. THESIS ADVISOR: ASST. PROF. SUPAVADEE ARAMVITH, Ph.D., 60 pp.

Today's scalable video coding technology provides a solution to multimedia content delivery over heterogeneous network to various kinds of terminals. In this thesis, the main focus is on the application of spatial scalability, which enables the delivery of a single multimedia content to users with devices supporting low video resolution to high resolution. However, the ubiquitous delivery network still uses a packet-based approach which cannot necessarily offer a guaranteed Quality of Service (QoS). Some losses may still be corrupt the video bitstream. For a highly compressed H.264 video, a single frame loss may have a dramatic effect on the visual quality of the affected frame as well as its succeeding frames.

Error concealment techniques are applied at the decoder side in order to estimate the lost information. Some works on error concealment for the H.264 video coding standard have been presented before, but their performance is found to be inadequate at times. In this thesis, we propose an improved error concealment scheme. We propose a motion estimation method which takes a weighted average of the motion vectors in a neighborhood of macroblocks from the previously decoded frame. Furthermore, the reconstructed frame is enhanced by filtering it with a hybrid median filter.

The proposed algorithm was found to have an average gain up to 0.8 dB over the state of the art BL_Skip which was also reflected the better quality in the subjective evaluation. Further investigation also shows that the proposed algorithm is able to suppress the effects of error propagation on the frames succeeding the lost frame.

Department.....Electrical Engineering…..... Student's Signature......................

Field of study.….Electrical Engineering... Advisor's Signature......................

Academic Year......2009…………

ACKNOWLEDGEMENTS

The author would like to thank his advisor, Asst. Prof. Supavadee Aramvith for her consistent efforts in helping to complete this thesis. Also, the author wishes to extend his gratitude towards Mr. Rhandley Cajote for his patience and inputs in discussing the H.264 video coding standard. Special thanks is expressed towards the ASEAN University Network/Southeast Asia Engineering Education Development Network (AUN/SEED-Net) and Japan International Cooperation Agency (JICA) for providing the necessary resources used in this study. Finally, the author would also like to thank all his family, colleagues in the Video Technology Group and friends who reminded him not to waver and keep pushing towards his final goal.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Objectives

The use of video in multimedia communications is increasingly becoming popular. With the deployment of new access technologies such as 3G and WiMax, more users can get access to packet delivered content, even if they are mobile. In turn, more devices are being developed also to display multimedia content such as video. Meanwhile, further development in broadband network technology enables the delivery of high bit-rate video streams. It can be noticed that the clients are heterogeneous in nature, in terms of connection speed and hardware capacities. In line with this, scalable video coding has been introduced. A scalable video is characterized by a single encoded video stream, but multiple substreams can be decoded from it [16]. Through the use of scalable video coding, it is possible to deliver a single video content but deduce subset videos with different frame rate, resolution and/or Signal to Noise Ratio (SNR), with the need to encode the original video content only once. On the distribution side, this means that only one encoded video needs to be stored instead of having multiple copies of the same video encoded with different qualities.

Despite the improvements in access technologies especially in terms of bandwidth capacity, it remains a challenge to guarantee a minimum Qualit of Service (QoS) as most communication networks are only able to provide best-effort services to its users. Transmitting sensitive data such as video over such a network will most likely introduce some packet errors and losses, resulting in degradation of video quality. Error concealment can be applied in the decoder in order to alleviate the damage caused by packet losses. The objective of this paper is to develop an error concealment algorithm for the scalable extension of the H.264/Advanced Video Coding (AVC) standard. However, this paper will also focus on the case of spatial scalability and whole frame losses.

1.2 Motivation and Significance of the Research Problem

1.2.1 Overview of Scalable Video Coding

The scalable video coding extension of the H.264/AVC standard supports multiple types of scalability including temporal scalability, SNR scalability, and spatial scalability. Temporal scalability describes a bit stream from which videos of different bit rates can be extracted. An SNR scalable video in the other hand supports different fidelity levels, often achieved by different levels of quantization at the encoder. It might be observed that temporal and SNR scalability are targeted at providing solutions for network heterogeneity and adaptation to different bit rates. Spatial scalability is used to represent video at different spatial resolutions. Spatial scalability is useful for delivering content at different resolutions [16]. For example, the full video resolution may be delivered to a high end computer, but only a reduced resolution version might be extracted for decoding in a mobile device. Conveniently, the use of a multiple layered bit stream is helpful in delivering a decodable stream of basic quality by applying more aggressive protection in chosen layers of the bit stream.

1.2.2 Prediction Tools in SVC

Inter-layer coding is a tool that aims to improve the coding efficiency of Scalable Video Coding (SVC) by exploiting statistical dependencies between layers. Similar to H.264/AVC, SVC also supports submacroblock level prediction to optimize the coding efficiency. Additional macroblock coding modes in SVC makes it possible to predict enhancement layer macroblocks from collocated macroblocks in the reference layer without transmitting additional side information. For example, in the case of spatial scalability, an intra-coded macroblock in the enhancement layer may be predicted from the upsampled version of the corresponding macroblock in the reference layer; this is called inter-layer intra prediction. Inter-layer inter prediction is generally similar to the inter-coding modes in H.264/AVC, however, the macroblock partitioning and motion prediction parameters are inherited from the base layer. This is also called inter-layer motion prediction. For the conventional inter-coded macroblock types of H.264/AVC, the motion vectors of the reference layer blocks can also be used instead of the usual spatial motion vector predictor. To further increase coding efficiency, another inter-layer prediction tool is introduced: inter-layer residual coding, which further reduces the bits required in transmitting the residual signals.

With the usage of residual prediction, the residual of the collocated reference layer blocks is subtracted from the enhancement layer residual and only the resulting difference is encoded using transform coding [16].

While coding efficiency is improved through the aforementioned prediction tools, it also renders the bit stream susceptible to errors and losses. On the other hand, it also demonstrates the high correlation existing among layers and neighboring frames.

1.2.3 Transmission of Scalable Video over Lossy Packet Networks

Unequal packet protection leaves the higher layers of the video stream vulnerable to packet losses. In some cases, an entire frame can be encapsulated into a single network packet; hence the loss of a single packet results into the loss of an entire video frame. This error will propagate to the succeeding frames which reference to this lost frame, degrading the quality of the video over an extended period. This provides the motivation to develop error concealment schemes that help recover or estimate the lost information.

Error concealment has been studied exhaustively in the H.264/AVC standard, but the scalable extension presents new challenges and remains an open research topic. There have been some interesting works for error concealment in SVC, some of which will be discussed in Section IV. One of these works was done by Chen et al. in 2006 [5] and has been adopted into the Joint Scalable Video Model (JSVM). These non-normative error concealment methods by Chen et al. will be further discussed also in Section IV. The error concealment methods integrated in the JSVM software provide basic means of recovering lost information at the decoder. We believe that exploiting more information and considering spatial and temporal redundancies will help in developing an improved error concealment technique.

1.3 Statement of the Problem

The latest video coding standard, H.264/AVC is capable of high compression by improved prediction and coding techniques. In some cases, an entire frame may be transported via a single packet. Such high compression makes the video stream extremely vulnerable to errors due to heavy use of inter prediction resulting in error propagation. In this worst case scenario, it is necessary to formulate an accurate estimate of the lost information, especially the motion vectors since they are needed in motion compensated prediction.

Error concealment often takes advantage of the high correlation of succeeding video frames. Hence the lost information may be inferred from the neighboring frames. In this paper, we want to address the problem of formulating an accurate estimate of the lost information in the case of whole frame losses in spatial scalable video.

CHAPTER II

BACKGROUND AND REVIEW OF RELATED LITERATURE

2.1 Scalable Video Coding

Scalable video coding can be applied as an alternative to simulcasting multiple copies of a video, each at different quality levels. However, the past attempts at implementing scalability in video coding were unsuccessful due to its low coding efficiency; i.e. the bitrate of MPEG-2 scalable coded video cannot match the bitrate of simulcasting its non-scalable equivalent. Also, scalable video coding implies the use of a more complex decoder. The latest video coding standard, H.264/AVC, has been extended also to support scalability features and is also known as H.264/SVC. H.264/SVC features an improved coding efficiency which makes it a viable solution for distributing of multimedia content to a heterogeneous client base [11]. However, an SVC decoder is more complex than a nonscalable decoder.

SVC is used to transmit bit streams which degrade gracefully as it adapts to varying network conditions. Target applications include video surveillance, erosion storage applications, video streaming and professional video editing and manipulating. Each SVC bit stream contains a basic quality substream which is backwards compatible to legacy H.264 decoders. Scalability provides progressive increase in decoded video quality by decoding the other substreams. This implies an increase in bit rate, possibly reducing the coding efficiency.

Since SVC is an extension of the current H.264 video coding standard, its design also incorporates a video coding layer (VCL) and a network abstraction layer (NAL). The encoded video content is found in the VCL, while the NAL packetizes the VCL units for storage and transmission. Not all NAL units contain VCL units as some NAL units are used to specify side information. A collection of NAL units that result in one decoded picture is called an access unit, and consecutive access units form a decodable portion of a video sequence. Scalability is implemented at the bit stream level; NAL units are simply dropped when there is a need to reduce the quality of the bit stream. Deciding which frames to drop is done with the information found in the NAL unit headers.

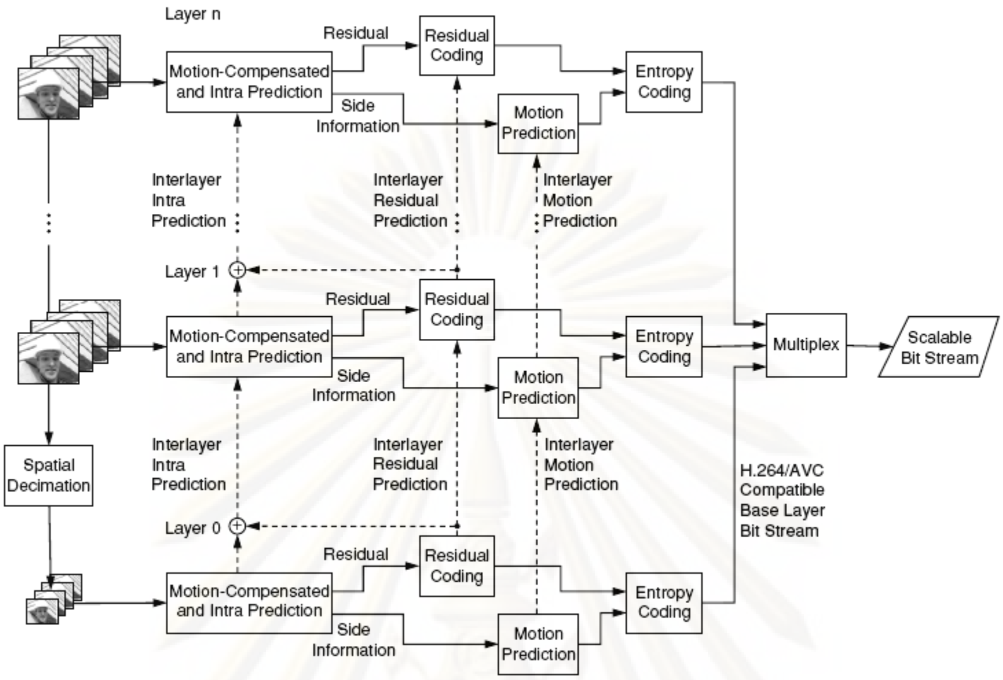


Figure 2.1. Typical scalable video encoder structure.

Figure 2.1 above shows a typical structure for an SVC encoder. As we can see, a scalable video bit stream can be comprised of multiple layers, each representing a frame in a particular resolution and fidelity. Layer 0, is called the base layer and is coded as a non-scalable bit stream compatible with non-scalable decoders. When a layer is used as a prediction source, it is called a reference layer. Layers which enhance the picture, whether be it in the spatial or temporal dimensions, or in fidelity, are called enhancement layers. SVC is reported to support up to 128 layers in a bit stream, but current implementations allow only for 47 enhancement layers, at most two of which are spatial enhancement layers.

Another highlight of SVC is that it only requires a single motion compensation loop for spatial and fidelity enhancement layer decoding. Only the intra-coded macroblocks and residual blocks from the specified reference layer needed in the interlayer prediction are reconstructed at the decoder, with their associated motion vectors. Hence, the computations for motion compensated prediction and deblocking are effectively reduced. Similarly, temporal scalability can be achieved by partitioning the access units into a temporal base layer and one or more temporal enhancement layers and restricting the encoding structure for each access unit of a specific temporal layer so that only access units of the same or a coarser temporal layer are employed for inter-picture prediction [16].

Multiple types of scalability are offered, namely Temporal Scalability, Spatial Scalability, and SNR Scalability. Each type of scalability can be applied independently or in combination.

Temporal Scalability translates into a bitstream where a video of lower frame rate can be extracted. This can be used to deliver video over a transmission channel with reduced bandwidth while retaining the resolution and fidelity of the decoded frames. The base layer of a temporally scalable video stream is encoded at the lowest frame rate which is desired to be supported. The base layer frames can be thought of as key frames which may be used to predict the additional frames brought by the enhancement layers.

Spatial Scalability on the other hand makes it possible to transmit video where substreams of lower resolution can be decoded. The application of spatial scalability is seen at distributing video content to terminals of differing computing power and resolution support. Videos of reduced resolution may be viewed by users on their mobile phone, while the full resolution may be enjoyed by users with larger and more powerful devices. The base layer in a spatially scalable video stream supports the lowest resolution desired. The enhancement layer is typically predicted from its corresponding base layer frame by using the upsampled base layer frame as a prediction source. This prediction method can also be called inter-layer prediction; inter-layer prediction methods have been developed to improve the coding efficiency of H.264/SVC.

SNR Scalability can be considered as a special case of spatial scalability. The base and enhancement layers are encoded in the same spatial resolution, but at different quantization parameters (QP). The base layer is encoded with a high QP value, giving it a lower quality. Visually, this typically translates into loss of texture, blockiness, and blobs.

2.2 Error Concealment in H.264/ AVC

Error concealment in H.264/AVC has been studied extensively over the past years. Error concealment algorithms often make use of spatial and temporal redundancies found in the received information at the decoder. Further refinements are introduced to ensure edge continuity in some regions or to reduce blocking artifacts. The reference software of the H.264/AVC Joint Model (JM) adopted two simple error concealment schemes: the Frame Copy and Motion Copy method. In the simplest terms, Frame Copy replicates the most recent reference frame used onto the lost frame, while Motion Copy replicates the motion information of the most recent reference frame and proceeds with motion compensated prediction. Many other works have followed and some interesting studies are reviewed in the following sections.

2.2.1 Pixel Based Motion Estimation Using Optical Flow

In 2005, Belfiore, et al. proposed an error concealment algorithm for dealing with whole frame losses in H.264/AVC [3]. They focused their study in the case of low-bit rate wireless video transmission. Their algorithm involves estimating the optical flow among several successive, previously decoded frames in order to estimate the lost motion information to predict the lost frame. Belfiore et al. pointed out that in their case, none of the previously implemented error concealment algorithms can be used to conceal the loss of an entire frame since at the time of their writing, earlier error concealment algorithms depended on spatiotemporal redundancies within the damaged frame i.e., some part of the frame is correctly received. Hence, they had to formulate an error concealment algorithm that utilizes other available information at the decoder. In their case, they exploited the frame buffer at the decoder to retrieve the motion and intensity information from available reference frames which were stored. Figure 2.2 below shows the flowchart of their proposed algorithm.

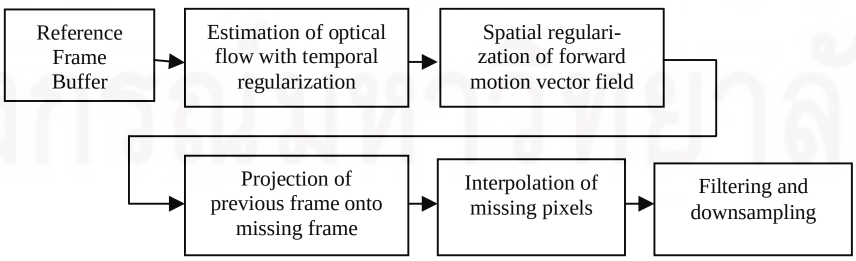


Figure 2.2. Pixel based MVE using optical flow

For each pixel in the lost frame F_t, its motion vector is estimated from the previous frame F_{t-1}. However, in order to ensure smoothness in motion, the motion vectors of F_{t-1} are re-estimated using information stored in the reference frame buffer; the estimated motion is constrained by the optical flow equation. Motion vectors are appropriately scaled depending on their temporal distance from F_{t-1}. Once the forward motion vectors have been estimated, the initial estimate for the motion field is regularized using a 7x7 median filter to compensate for pixels whose forward motion vector is unknown; this is especially true in high motion sequences where some macroblocks may be encoded in intra mode. Further spatial smoothing is performed using a 2-dimensional 16x16 median filter. Half-pixel resolution is used to refine the estimated motion vectors.

The lost frame is first reconstructed at half pixel resolution (i.e., twice the number of samples in the horizontal and vertical directions). The pixels of the lost frame are projected from an upsampled version of the previous frame. When multiple projections of pixels have overlapped on the reconstructed frame, the average of the overlapping pixel values is used. The reconstructed frame is then filtered with a 9x9 median filter to account for any missing pixels. Finally, the reconstructed frame is downsampled to its correct size.

The algorithm is found to produce good objective and subjective results since the motion is well smoothed. However, other papers citing this algorithm often speak of its rather high computation count due to multiple median filtering operations.

2.2.2 Block Based Motion Estimation Using Optical Flow

In an effort to reduce the computations involved for error concealment, Baccichet et al. proposed a block based approach to motion estimation [3]. Their algorithm retains the use of optical flow for the estimation of motion vectors, but most of the filtering operations are performed on the block level, reducing the required kernel size, as well as the number of computations compared to Belfiore's algorithm.

Support of multiple reference frames in H.264/AVC is used in order to find the most suitable reference frame for the concealment process. From the reference picture buffer, previously decoded frames are inspected to see which one among the temporally nearby frames has the highest count of inter coded macroblocks. The motion vectors on the chosen reference frame are projected onto the lost frame and are scaled according to the temporal distance between the reference frame and the lost frame.

The motion field of the lost frame is estimated twice: once at the macroblock level, and once more at the block level. The goal of macroblock level motion field estimation is to determine the image regions where generally uniform motion exists.

After all macroblocks have been checked, it may be possible that some of the macroblocks remain empty. The empty regions are reconstructed using a similar condition above, except that 4x4 block size is used instead of 16x16. This step is iterated until all blocks have been filled. Finally, the lost frame is reconstructed using motion compensated prediction. As expected, there is a trade-off between computational complexity and concealment performance.

2.2.3 Frame Error Concealment Technique Using Adaptive Inter Mode Estimation

This paper takes a two-step approach to the problem of error concealment. The first step involves bidirectional motion vector extrapolation, and second, adaptive inter-mode estimation. Bidirectional prediction may be helpful especially in cases where object occlusion occurs. The complexity of their proposed algorithm is low enough to be implemented in real-time [10].

In the concealment process, motion vectors from both past and future frames are considered. Motion estimation is performed using block size of 4x4.

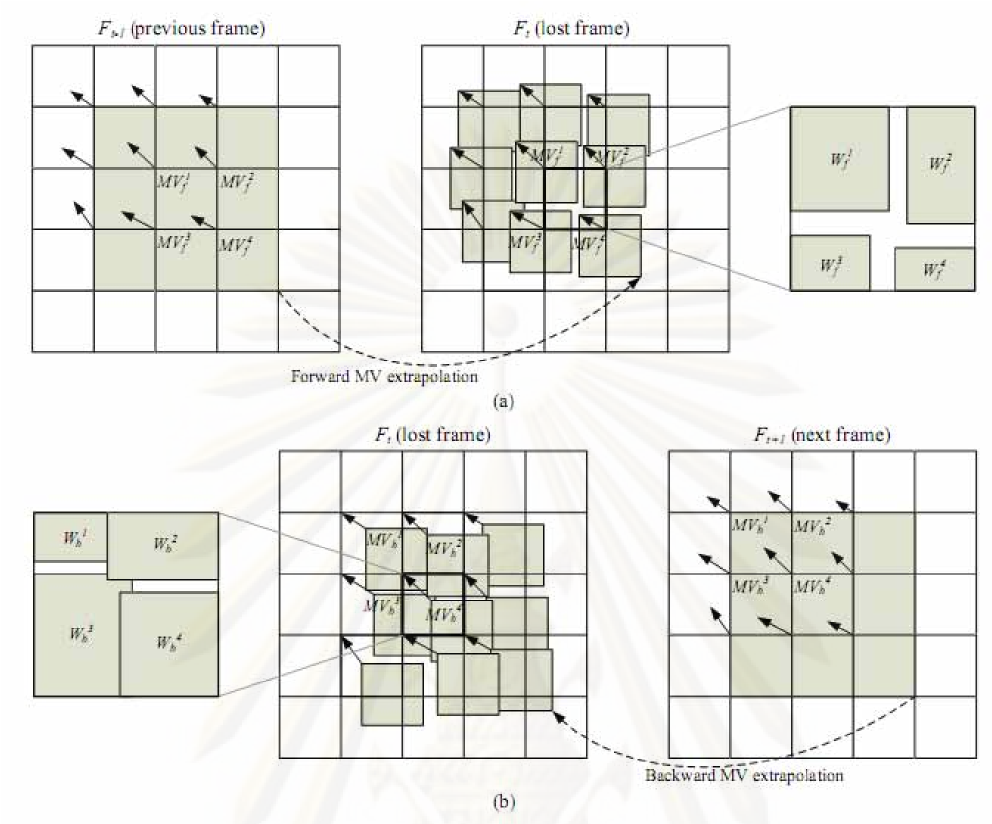


Figure 2.3. Bidirectional motion vector projection in H.264/AVC

Figure 2.3 above shows how the motion vectors are projected from the previous frame and from the future frame. Prediction from the previous frame is also called forward prediction since the prediction direction goes forward in time; opposite is the case for backward prediction or prediction from the future frame. In contrast to the conventional motion copy error concealment method integrated within the JM reference software, the method in [10] considers all motion vectors which affect the currently concealed macroblock by checking which projections overlap onto the macroblock. The enlarged portion in Figure 4.4 above illustrates how four blocks from the previous frame F_{t-1} are extrapolated onto the lost frame F_t using the MVs from the previous frame. Also, these extrapolated blocks are found to have overlapping regions on the macroblock to be concealed shown as the block with bold outline. The forward motion vector for the target macroblock is estimated by equation (1).

$$M\hat{V}_f = \frac{\sum_{i=1}^{N} W^i{}_f \cdot MV^i{}_f}{\sum_{i=1}^{N} W^i{}_f} \qquad (1)$$

Where W_f is a weighting factor corresponding to the number of pixels in the overlapping area and N is the number of overlapping blocks. Backward motion vector estimation is performed in a similar manner, except that the direction is reversed. The backward motion vector estimate is calculated using equation (2):

$$\hat{MV_b} = \frac{\sum_{i=1}^{N} W^i{}_b \cdot MV^i{}_b}{\sum_{i=1}^{N} W^i{}_b} \tag{2}$$

In an earlier work of the authors, the final motion vector estimate is taken as the average of the forward and backward motion vector estimates. However, artifacts may arise especially if the forward and backward motion estimates have differing directions. A new approach is taken by taking into consideration the similarity in direction of the forward and backward motion estimates and correlating it with the size of the overlapped regions. An optimal pair (k,l) of forward and backward motion vectors is found by finding the maximum overlapped area and minimum difference between the motion vectors as shown in equation (3).

$$(k,l) = \arg\max\left(\left(W^i{}_f + W^i{}_b\right) - \alpha\left|MV^i{}_f - MV^i{}_b\right|\right) \tag{3}$$

Where α is a factor controlling the influence of MV difference. A suitable value for α was experimentally found and fixed to 0.6.

After the motion vectors for the lost frame have been estimated, it is still possible that some image regions have no overlapping blocks projected onto them, and in addition, blocking artifacts may be seen along the block edges. In order to address this issue, the inter-mode of each 16x16 macroblock in the lost frame is estimated adaptively.

It is assumed that all macroblocks in the lost frame is encoded using one of the inter modes of H.264/AVC. Figure 2.4 below illustrates how the inter modes are estimated from the 4x4 blocks.

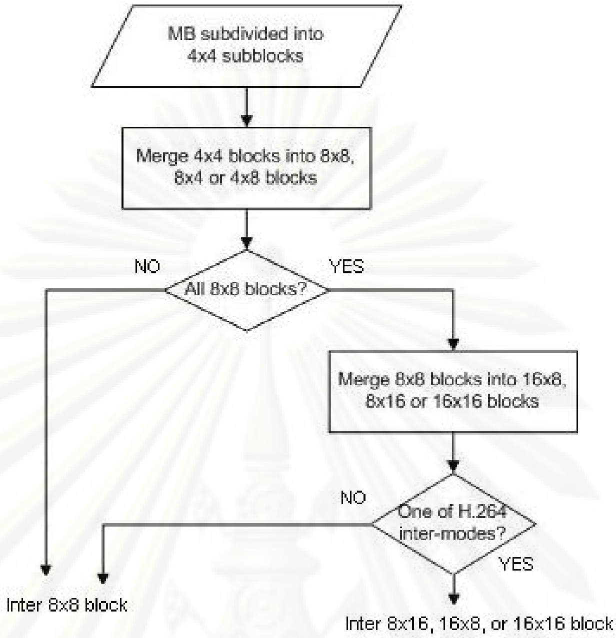


Figure 2.4. Adaptive inter-mode estimation

Neighboring 4x4 blocks are merged depending on the correlation of their motion vectors. The degree of similarity is defined as a threshold in MV difference. The purpose of inter mode estimation is to assign motion vectors to the hole block regions where no block has been projected on. The process of inter mode estimation is of fairly low complexity and was found to be effective.

2.2.4 Summary of Error Concealment in H.264/AVC

In summary, error concealment algorithms in H.264/AVC have distinguishing characteristics, each with their own highlights and weaknesses

Pixel-based error concealment algorithms – these algorithms perform operations on a pixel-wise or sub-pixel basis. For example, the estimated pixel in the frame to be concealed may be taken as a weighted average of neighboring pixels in its reference frame. Another case is when motion estimation is done per pixel. Pixel-based algorithms generally yield better precision and smoother gradients, at the expense, however of greater number of computations.

Block-based error concealment algorithms – in contrast to pixel-based algorithms, block-based error concealment algorithms tend to produce less refined outputs and are prone to blocking artifacts. However, considerable savings in computational resources can be achieved, which is desirable for real-time implementation. One may opt to use a smaller block size in order to achieve a more refined output at the expense of some increase in computation. Blocking artifacts resulting from mismatched edges and discontinuities may be alleviated by applying a deblocking filter to the reconstructed frame.

Multiframe error concealment algorithms – these algorithms use more than one reference frame as information sources. Most of these algorithms rely on the assumption that the content between successive video frames is differential and changes smoothly. One special case of multiframe error concealment involves bidirectional prediction. One of the source frames is a frame that precedes the lost frame in display order, and another source frame is a frame that comes after the lost frame. Other works have shown that using a bidirectional approach in motion estimation gives a more accurate motion estimate.

2.3 Error Concealment in H.264/SVC

Although H.264/SVC is in many ways similar to H.264/AVC, the problem formulation for error concealment is quite different. Recall that H.264/SVC uses a multilayered video stream and the use of inter-layer prediction is possible, and different types of scalability can be implemented. In this paper, however, we will focus on the case of spatial scalability.

2.3.1 Non-normative Error Concealment

In 2006, Chen et al. worked on error concealment for SVC and developed several algorithms to tackle frame losses in a spatial scalable bit stream. Later, their work was recognized and adopted as non-normative error concealment methods for the JSVM reference software [5].

This paper presents categorizes error concealment methods into “inter layer error concealment’ and “intra layer error concealment” depending on which layers are designated as concealment source and destination. Two intra layer error concealment methods are proposed namely Frame Copy and Temporal Direct error concealment methods. These methods operate on a single layer without requiring additional information from other layers.

Frame Copy (FC) is an intra layer concealment method; it uses the first frame in the Reference Picture List 0 (RefPicList0) as a concealment source where it copies the pixel values from the reference frame onto the lost frame.

Temporal Direct (TD) Motion Vector Generation is also an intra layer error concealment algorithm; it assumes that the missing macroblocks were encoded using the temporal direct mode where the reference indices and motion vectors are estimated. For a lost enhancement layer frame, the motion vector for a lost macroblock is estimated from its collocated MB in its reference picture. The motion vector is scaled according to its temporal distance from the concealed frame.

The two remaining methods are inter-layer error concealment methods, which in contrary to the two prior methods, make use of information across different layers. These methods are the Motion and Residual Upsampling method, otherwise known as BL_Skip and Reconstruction Base Layer Upsampling (RU).

BL_Skip uses information from both the base layer and enhancement layer and is therefore considered as an inter layer error concealment algorithm. The residual of the base layer is upsampled for use at the enhancement layer; however, the motion compensation is done within the enhancement layer using upsampled motion fields. The method can immediately be applied to the enhancement layer if the base layer is correctly received. If the base layer is lost, it is first concealed using the TD method before performing BL_Skip.

Reconstruction Base Layer Upsampling (RU) is an inter layer error concealment algorithm; the lost enhancement frame is concealed by upsampling the corresponding base layer frame with the H.264/AVC 6-tap filter. If the base layer frame is lost, it is first concealed using FC prior to RU. Lost key frames are also concealed using FC. The necessary Memory Management Control Operation (MMCO) and Reference Picture List Reordering (RPLR) commands are executed in the concealment process to emulate the reference and non-reference picture patterns at the encoder.

Experiments show that he BL_Skip method consistently yields the best concealment results among the four methods, both in objective and subjective evaluation, although occasional blocking artifacts appear worse compared to RU. However, the BL_Skip method still holds an upper hand when it comes to computational complexity. RU requires multiple-loop decoding since the reference layer has to be reconstructed first before concealment can be performed on any higher-layer frames. BL_Skip however takes advantage of the single loop decoding feature I H.264/SVC. Temporal direct mode was found to be effective in image regions which are quasi static and has the advantage of preserving texture better when compared to RU method. Frame copy performs the worst, but is the simplest to implement. BL_Skip generally performs better than RU since motion compensated prediction is performed at the enhancement layer. Also, the upsampling process in the RU method also tends to produce blurry images.

2.3.2 Frame Error Concealment Technique using Hallucination

The main drawback of the BL_Skip method is its blurry output which is mainly attributed to the upsampling process used for predicting the residual. In relation to this, Ma Q. et al. proposed to use a technique called hallucination in order to conceal lost enhancement layer frames from the base layer. Hallucination is a super resolution technique used to generate a high resolution image from a single low resolution image sample with a pre-trained database. The training images used come from previously decoded frames stored in the database. The effectiveness of the concealment algorithm depends on the correlation of consecutive frames in a video sequence [12].

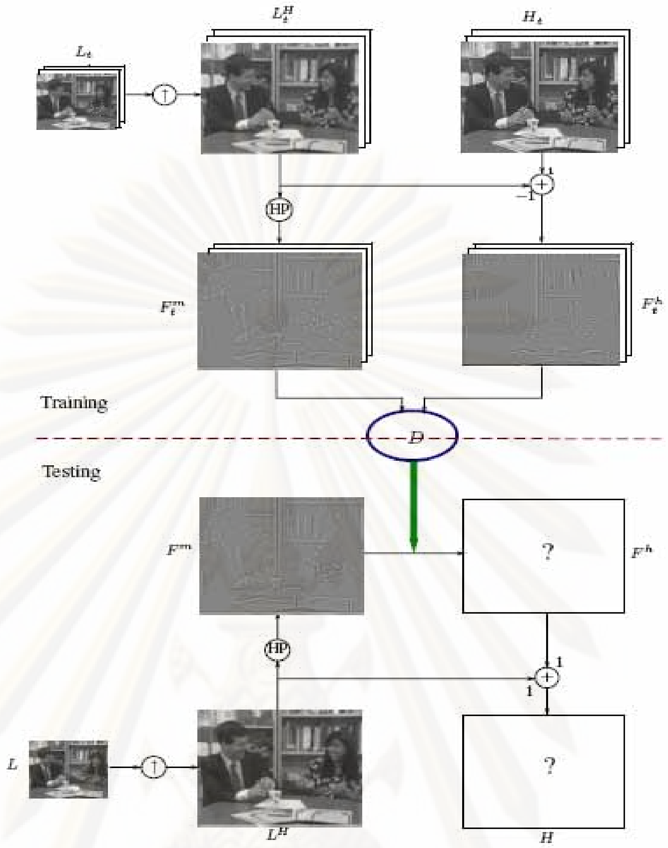


Figure 2.5. Error concealment using hallucination.

Figure 2.5 shows the general process of hallucination. Training images are generated by downsampling the high resolution images. The low and high resolution images are paired { $L_t - H_t$, t = 1,…,P} and input to the system. The low resolution image L_t is upsampled to L_t^H which has the same spatial resolution as H_t. The high frequency component F_t^H is extracted by taking the difference $F_t^H = H_t - L_t^H$. Meanwhile, the medium frequency component F_t^m is extracted by high pass filtering of L_t^H. F_t^m and F_t^H are normalized by an energy map and co-located patch pairs are extracted from them. The collected patches are stored in a database D of patch pairs $\{p^m - p^h\}$. The size of p^m is equal to or larger than p^h but is typically set to 7x7 in most researches.

The reconstruction process mirrors the training process, except that the high resolution image H is not present, as this is what the concealment algorithm aims to reconstruct from a low resolution image L. Let LH be the upsample of L and Fm be the high pass filtering of LH . Example-based single-image super-resolution technique assumes that the high frequency component Fh is only dependent on the medium frequency component Fm. Using Fm and the database D, we can predict Fh. For a patch qm in Fm, we find the most similar mid frequency patch pm in D, and consequently its pair ph. We add the high frequency patch ph to Fm to generate the high frequency component Fh. In cases when there are many possible matches to qm, the problem is formulated as a Markov Random Field and the best combination of the candidates is computed.

For a lost frame in the enhancement layer, it is hallucinated using its corresponding base layer frame in order to conceal it. The hallucination database is trained using high resolution and low resolution frames in the temporal neighborhood. As such, due to the high correlation of temporally adjacent frames, hallucination is expected to effectively conceal the lost enhancement frame.

There are two ways by which hallucination can be used for error concealment. The first approach is performed out of the decoding loop (i.e. post processing). Erroneous frames can be concealed with hallucinated ones. Because it is performed after the decoding process, this out of loop concealment may introduce some delay. Another concern is about error drift; there may be more error frames than those frames which only suffer from enhancement information loss. The decoder must decide which frames to hallucinate and which ones will not. The second approach includes the hallucination process in the decoding loop. Hallucination-concealed frames can be used as future reference frames. The superior performance of hallucination over BLSkip makes it better in terms of stopping error drift. After experiments, it is also found that in loop hallucination incurs less delay and performs better than out-of-loop hallucination.

As mentioned before, the training images are sourced from nearby temporal neighbors; particularly in this paper, two previously decoded frames are used. In case of hierarchical-B coding structure, one of the training pictures is chosen as the nearest, correctly received I-frame prior to the lost frame, and the other is the I-frame after the lost frame. The hierarchical-B coding structure allows for the required I-frames to be decoded and reconstructed before attempting to decode the damaged B frame. For the IPPP coding structure, the two most recent, correctly received previous frames are used as training images.

Only the luma component is concealed by means of hallucination. The chroma components are simply upsampled from the base layer using bilinear filtering.

The algorithm is tested to conceal errors in the enhancement layer, and the base layer is assumed to be received correctly. An average gain of about 2.5 dB was reported when compared to the BL_Skip method. It also successfully overcomes the BL_Skip method in terms of visual clarity since it is able to produce sharper images. However, the hallucination technique demands computational resources for the database and patch retrieval. Also, the method is designed primarily for the concealment of enhancement frame losses only; no method is prescribed for concealing base layer losses.

2.3.3 Error Concealment for SNR Scalable Video Coding

M. M. Ghandi and M. Ghanbari also proposed an error concealment algorithm in 2005, although their considered SNR scalability instead of spatial scalability. On the other hand, SNR scalability can be considered as a special case of spatial scalability where the base layer and enhancement layer have the same spatial resolution but differing quality levels [8].

In a non-scalable decoder such as H.264/AVC, erroneous macroblocks can be concealed using its neighboring macroblocks or even macroblocks from the previous frames. It is rather simple since an H.264/AVC bitstream has only one layer.

When an MB enhancement layer is corrupted, it can be concealed from the previous enhanced frame(s) or the current base layer frame. Since the base layer is often more protected compared to the enhancement layers, it is deemed a more reliable concealment source. Prediction methods using the information of the base layer is also described as an *upward method* since the prediction goes from a lower layer to a higher layer. On the other hand, if the enhanced reference frame is received without error, it can be a better quality concealment source. Previous works carried out offer mode decision algorithms to choose between the base layer and enhanced layer frames, as well as a combination of both. However, there are no works heavily exploiting the available motion vector information.

[8] focuses on the error concealment in the enhancement layer where two scenarios are observed. In the first scenario, the enhancement layer MB is correctly received, but the base layer is in error; instead of discarding the enhancement layer

information, it is considered in the concealment process. In the second scenario, the base layer is received correctly, while some macroblocks in the enhancement layer have been corrupted. Concealment through motion compensation is carried out, with the candidates collected from the base layer and available enhancement layer information. The concealment algorithm uses picture continuity along the block edges as criteria, together with compatibility of coefficients with the base layer. The proposed algorithm is compared with conventional error concealment method.

In order to design an effective error concealment technique at the decoder, it is necessary to study the prediction schemes used in the encoding process. In this SNR scalable method, there are three prediction modes described in Figure 2.6 below.

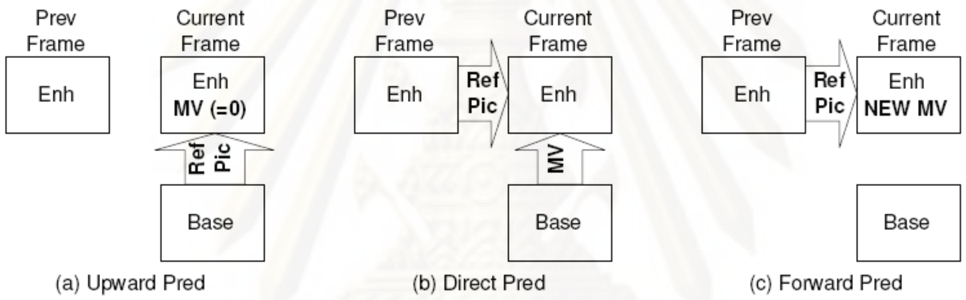


Figure 2.6. Prediction modes in SVC

The first case depicts the *upward* prediction mode; the current base layer frame is used as a reference for the current enhancement frame, and the motion vectors are set to zero. In the *direct* mode, the previous enhancement frame is used as a reference, while using the motion vectors of the base frame. Finally, the *forward* prediction mode involves generating a new set of motion vectors from the previous enhancement frame. After prediction, the residual between the predicted frame and the source is taken, transformed, quantized, and entropy coded. The quantizer for the enhancement layer (QE) is larger than that for the base layer (QB) in order that the quality is improved upon decoding the enhancement layer.

In a scalable video codec, the data for an individual MB is split over several packets. When packet losses occur, some of the MB info may be lost. A table of various possible scenarios is presented in Table 2.1 below.

Table 2.1. Prediction mode for different loss scenarios.

Situation	**Original Enhancement layer prediction mode**	**Concealment Candidates**
Enhancement layer lost	Unknown	Upward, Direct
Base layer lost	Upward/Direct	Upward, Direct, Decoded
	Forward	Decoded

The first situation suffers from the loss of the enhancement layer. At the decoder, it cannot be known how the lost enhancement MB was predicted in the encoder. Being aware that the base layer is typically more protected and reliable than the enhancement layers, the concealment candidates are then taken using the upward and direct prediction modes.

The next situation represents the inverse case of the first; if the enhancement MB was encoded using either upward or direct mode, it becomes useless since it was predicted from the base layer, which happens to be corrupted. Standard decoders will normally discard the enhancement MB in such a case. For the proposed method, the enhancement layer is still used, but with a reconstructed base layer MB. This is termed as “Decoded” in the above table.

For the concealment process, we consider at most three candidates: upward concealment, direct concealment, and the decoded data. From the candidates, it is desired to choose the one which yields the minimum distortion. In practice, a discontinuity measure is used instead. The concealment discontinuity (D) consists of two components D_b and D_e. The former represents the candidate’s compliance with base layer coefficients and the latter describes the discontinuity at the enhancement layer boundaries.

D_e is calculated as the absolute difference between the edge pixels of the candidate and its neighboring blocks assuming they are available. The normalized equation is shown below in equation (4).

$$D_e = \frac{1}{N}\sum_{i=0}^{N-1}\left|c_i - n_i\right| \tag{4}$$

D_b on the other hand, is calculated using the base layer data. Since SNR scalability is used, it is safe to say that the enhancement layer actually contains the quantization distortion of the base layer, in the form of residual data. This possible since the enhancement layer was quantized using a QE instead of QB. It is possible that the quantization error will not be equal to zero, but its value should approach zero. The coefficients (R_i) for incorrect candidates are used as a criteria for calculating D_b.

$$D_b = \frac{1}{256}\sum_{i=0}^{255}\left|R_i\right| \tag{5}$$

The value of D_b, calculated by equation (5), is only if the base layer is not lost or corrupted. Ultimately, we choose the candidate that gives the lowest overall discontinuity measure (D).

$$D_{MIN} = \{(D_b + D_e)/2, D_b or D_e, None\}$$

The first case applies when both D_b and D_e exist, while the second condition is for only one of them is present, and lastly, when neither is available. In case of the third condition, upward concealment is performed for that MB.

The proposed algorithm was tested against the upward concealment method. In the first test scenario, the base layer and enhancement layer are assumed to be protected equally and subject to the same bit error rate. Multiple test runs were performed and the average Peak Signal to Noise Ratio (PSNR) was taken. The maximum gain of the proposed method was posted at 3 dB.

The second test scenario assumes unequal error protection; specifically, the test assumes that the base layer is received successfully without errors. The proposed algorithm still outperforms the conventional upward method, although the improvement this time is marginal. This is due to the additional information from the enhancement layer used by the proposed method, while the conventional method only relies on base layer information.

In general, all error concealment algorithms attempt to reconstruct the lost frame from information available at the decoder. This paper aims to solve the problem of whole frame losses in a scalable video bit stream. As such, the motion information as well as the residual information of the current frame are both not available at the decoder. The motion in consecutive video frames is assumed to have high correlation, and this correlation decreases as temporal distance between frames increases. With this in mind, we wish to utilize the motion information from nearby frames in order to estimate the lost motion information.

2.3.4 Comparison of Error Concealment Methods

We summarize in Table 2.2 the pros and cons among selected error concealment algorithms reviewed above. In the next section, we discuss the performance of the chosen benchmark algorithm, the BL_Skip, which is regarded as the best among the non-normative error concealment methods.

Table 2.2. Summary of Error Concealment Methods

Concealment Method	**Advantage**	**Disadvantage**
Frame Copy	• Least computationally complex	• Performs poorly in high motion sequences • Causes jerkiness in playback
Temporal Direct	• Performs well on low motion sequences	• Derived motion vectors may not reflect true motion vectors in high motion sequences
Reconstruction Base Layer Upsampling	• Less chances of blocking artifacts	• Requires multiple loop decoding • Blurry output due to upsampling process
BL Skip	• Requires only single loop decoding	• Effective only if base layer information is present • Blurry output due to use of upsampled information
Frame Error Concealment Technique using Adaptive Inter-mode Estimation	• Bidirectional approach yields good motion vector estimate • Inter-mode estimation compensates for hole-block problem	• No implementation in JSVM • High complexity
Frame Error Concealment Technique using Hallucination	• Output sharper compared to BL Skip • Performs well in low motion sequences	• Cannot recover from base layer loss • High complexity • Difficulty in concealing high motion sequences

2.3.5 Simulation on BL_Skip Performance

Among the error concealment methods for SVC, the most widely considered for comparison is the BL_Skip method. It is regarded as the best among Chen’s non-normative error concealment methods. We tested the performance of the BL_Skip algorithm in various packet loss conditions. We observed that the performance of BL_Skip is reasonably acceptable in the condition that the enhancement frame is lost while the base layer is present. However, when both the base layer and enhancement layers are lost, the BL_Skip method reverts to recovering the base layer motion information through the temporal direct method. The estimated motion information is then scaled accordingly for the enhancement layer frame. The resulting concealed frame often exhibits blocking artifacts resulting from the poor motion estimate. The situation reflects that the estimated motion field is dissimilar to the true motion field. Another observation is that the concealed frame sequence exhibits artifacts similar to the Discrete Cosine Transform (DCT) basis functions and salt and pepper noise. Due to the use of predictive coding techniques, the artifacts are often carried on to the frames succeeding the damaged frame, causing the overall quality to drop.

CHAPTER III

METHODOLOGY

3.1 Proposed Method

Through performing tests, we have identified the weakness of the BL_Skip method, which mainly is its diminishing performance in the absence of base layer information. We propose an improved motion estimation method that can reduce the blocking artifacts and produce a more accurate motion estimate. Post processing in the pixel domain can be implemented in order to reduce the unwanted artifacts in the concealed frame.

Several works on error concealment have been proposed for H.264/AVC. Meanwhile, error concealment in H.264/SVC has yet to be fully explored especially since several scalability options are available. However, different scalability options may require different approaches in error concealment.

In H.264/AVC, several works we reviewed use the motion information extracted from the previously decoded frames. With the aim to improve the quality of enhancement over conventional methods, spatiotemporal redundancies in the received video stream are exploited. Belfiore and Baccichet both proposed methods to estimate the true motion in the damaged frame. The motion field is estimated over the past few frames and subject to the optical flow equation. Both papers work around the assumption that the motion among successive frames are highly correlated or changes very little. On the other hand, [10] takes a bidirectional approach in estimation of motion vectors.

One common denominator of these motion estimation-based concealment algorithms is the problem of hole-blocks following the motion compensated prediction process. It is possible that some regions in the frame do not have any blocks or pixels projected from the reference frame. Hence, simply estimating the lost motion vectors is inadequate to conceal the damaged macroblocks. Belfiore et al. approaches the problem by interpolating the missing pixels using a 9x9 median filter. Meanwhile, Baccichet et al. uses statistics collected from each macroblock to determine the best motion vector candidate for it. The initial pass is performed for 16x16 macroblocks; in case some macroblocks have empty motion fields, a second pass is performed using a smaller block size and lower thresholds. The process is iterated until all blocks have been filled. Meanwhile, [10] takes an interesting approach since it attempts to estimate the inter mode in which the lost block was encoded. The motion field is initially estimated at the 4x4 block level. In compliance with the possible configurations for inter-mode encoding, neighboring 4x4 blocks are checked to see if their motion vectors are similar. Upon satisfying a certain threshold, neighboring blocks can be merged. This merging process enables unfilled blocks to be assigned motion vectors similar to its neighbors.

With scalable video coding, one has more options than extracting information from spatiotemporal redundancies within the video sequence. One can also infer information from another layer such as motion vectors, quantized coefficients, and even residual information. Recall that this paper focuses on the case of spatial scalability. The base and enhancement layers are encoded at different spatial resolutions, with the base layer having lower spatial resolution. It is permissible to formulate predictions from base layer to enhancement layer, provided that the appropriate scaling is performed.

The accuracy of the estimated motion vectors has a great influence on the quality of concealment since the decoding process involves motion compensated prediction. Even with works on H.264/AVC, the trivial Motion Copy method performs inadequately, often exhibiting blocking artifacts and discontinuities along block edges. As such, we wish to use a more accurate set of motion vectors. We propose a method of estimating the motion vectors for a lost enhancement layer frame from the motion information gathered in the previous frame.

3.1.1 Proposed Motion Estimation Scheme

The motion vector for a lost macroblock is calculated via a weighted estimate of the motion vectors in the neighborhood of the macroblock. However, the qualified neighbors to be considered are those whose projections overlap with the macroblock to be concealed. Weights are assigned to the neighboring motion vectors based on the area in pixels that overlaps on the current macroblock location. The motion estimation is carried out at the base layer resolution and upsampled for use in the enhancement layer.

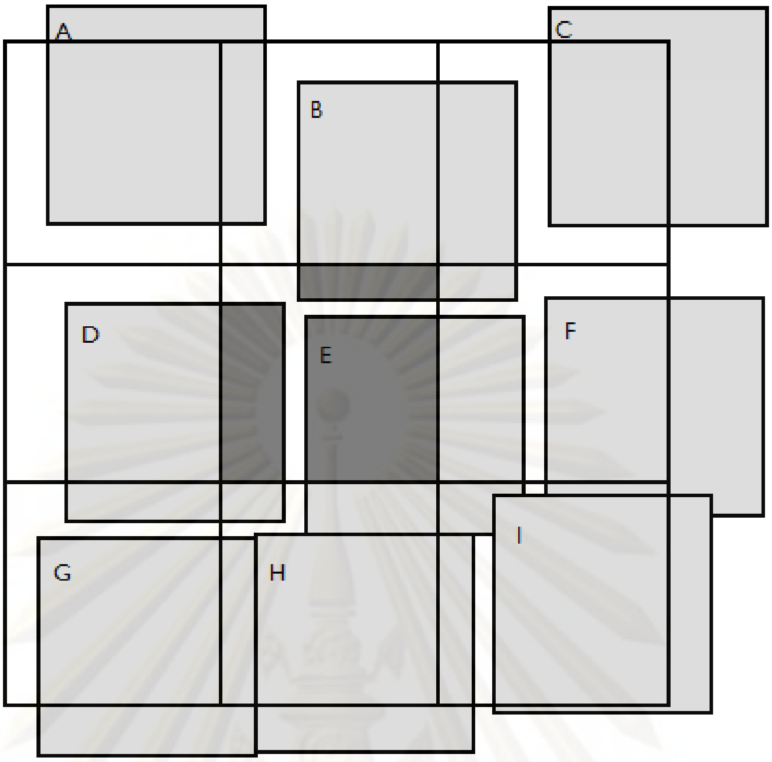


Figure 3.1. Macroblock projections and overlapping regions.

In Figure 3.1 above, let the center macroblock be a lost macroblock which needs to be concealed in the current frame. Macroblocks A to I represent the projections of the macroblocks in the previous frame onto the current frame. Macroblocks B, D, and E clearly overlap with the macroblock to be concealed; the overlapped area is marked as the dark gray area. Therefore the motion vectors of B, D, and E will be considered for estimating the motion vector of the lost macroblock. In general, the motion vector estimate is computed by equation (6):

$$\widehat{MV} = \frac{\sum_{i=0}^{N} MV_i \cdot W_i}{\sum_{i=0}^{N} W_i} \tag{6}$$

Where N is the number of overlapping macroblocks, and W is a weighting factor corresponding to the number of pixels in the overlapping region. The worst case we consider for loss is the loss of the base layer frame; even if the enhancement layer frame was correctly received, since it relies on its corresponding base layer for prediction, it becomes practically useless [20]. Hence, we perform the motion estimation at the base layer using a block size of 8x8. The estimated motion field is then upsampled for motion compensated prediction at the enhancement layer.

3.1.2 Enhancement of Reconstructed Frame

The artifacts present in the concealed frame after BL_Skip have similar characteristics with impulse noise. In order to suppress these artifacts, we apply a hybrid median filter to the reconstructed frame. Median filters are known for their edge preserving characteristics. The hybrid median filter by Garcia [7] was designed to have better edge preserving characteristics than a square kernel median filter. It uses an N x N box and performs a three-step ranking operation independently on three spatial directions. Fig. 3.2 below shows a 5 x 5 hybrid median filter.

$$\begin{bmatrix} D & * & R & * & D \\ * & D & R & D & * \\ R & R & DCR & R & R \\ * & D & R & D & * \\ D & * & R & * & D \end{bmatrix}$$

Fig 3.2. 5x5 Hybrid Median Filter by Garcia.

Three median values are computed; for the horizontal and vertical directions, the median MR is taken from the R pixels. For the median MD in the diagonal direction, the D pixels are considered. The filtered median value is taken as the median of MR, MD and the central pixel C. The overall process is described below in Figure 3.3.

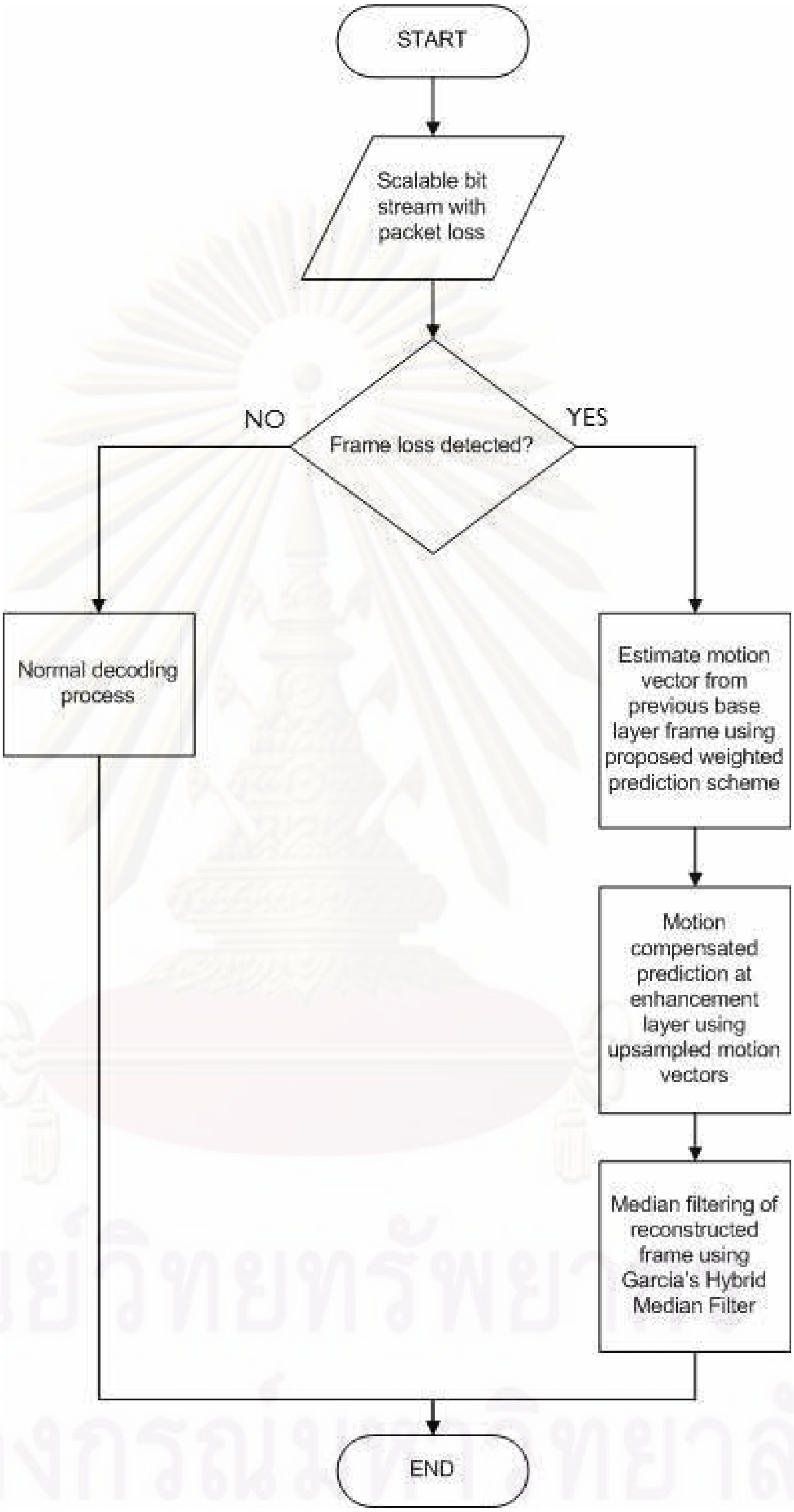


Fig. 3.3. General process of the proposed error concealment.

The proposed method may be viewed as an improvement to the BL_SKIP algorithm. As we may recall, in the case of base layer loss, the BL_SKIP algorithm reverts to the Temporal Direct motion vector generation method in order to estimate the lost frames. Experimentally, we have observed that the performance of BL_SKIP degrades drastically in the case of base layer losses, which leads us to conclude that the TD method is not a satisfactory method of estimating the lost motion information. At this point, we introduce our proposed motion estimation scheme which calculates the motion vector from the previous base layer frame. The calculated motion vectors are then upsampled in proportion to the resolution of the enhancement layer frame.

Upon experimenting with the proposed motion estimation scheme, some noisy artifacts were still observed in the reconstructed sequence, leading to the proposed application of spatial filtering using a median filter. The median filter is chosen because of its edge preserving characteristics which is a desired feature in order to reduce the effect of blurring on noise reduction operations. A number of median filters were tested, but Garcia's [7] hybrid median filter was found to be most effective experimentally.

CHAPTER IV

EXPERIMENTS AND RESULTS

4.1 Experiments and Results

The effectiveness of the proposed method is compared against the conventional BL_Skip algorithm in packet loss scenarios. Two scenarios are tested: enhancement layer frame loss only and loss of both enhancement and base layer frame. The tests are carried out using Joint Scalable Video Model (JSVM) 8.5 reference software. Multiple test video sequences were used and encoded in two-layer spatial scalability. The base layer is encoded in QCIF resolution or 176x144 pixels while the enhancement layer is encoded in CIF resolution or 352x288, both at 30 frames per second. An IPPP coding sequence is used. Constant quantization parameter (QP) encoding is also applied, but we test for QP values of 28, 34 and 40, which reflect medium and low bit rate transmission respectively. A packet loss rate of 5% is simulated for each layer.

A lost frame is detected by inspecting the frame number of the incoming frame. Logically speaking, successive frames would simply increment the count of the frame number from what was previously received. When the next frame arrives and its frame number

The first set of tests were done with the base layer intact, hence we can independently evaluate the effectiveness of the median filtering scheme. In our experiments, we varied the size of the median filtering window and determined that a window size of 3x3 was suitable for most concealment conditions. The window size was kept at that size for the rest of the experiments.

Table 4.1. Average PSNR for concealed sequences with 5% loss in enhancement layer, QP = 28.

Sequence (QP28)	**BL_Skip**	**Proposed**
Coastguard	**34.26 dB**	32.12 dB
Container	**35.32 dB**	31.85 dB
Crew	32.79 dB	**33.71 dB**
Football	**35.64 dB**	34.64 dB
Foreman	29.23 dB	**30.23 dB**
Mobile	19.80 dB	**20.67 dB**
News	**34.37 dB**	33.58 dB
Soccer	**34.89 dB**	33.55 dB

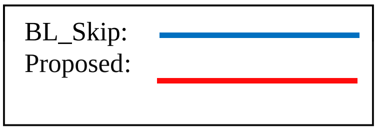
Table 4.2. Average PSNR for concealed sequences with 5% loss in enhancement layer, QP = 34.

Sequence (QP34)	**BL_Skip**	**Proposed**
Coastguard	**30.23 dB**	29.65 dB
Container	**31.17 dB**	29.94 dB
Crew	30.03 dB	**31.02 dB**
Football	31.62 dB	**31.65 dB**
Foreman	28.22 dB	**29.44 dB**
Mobile	19.59 dB	**20.73 dB**
News	**31.71 dB**	31.62 dB
Soccer	**31.04 dB**	30.89 dB

Table 4.3. Average PSNR for concealed sequences with 5% loss in enhancement layer, QP = 40.

Sequence (QP40)	**BL_Skip**	**Proposed**
Coastguard	**26.61 dB**	26.59 dB
Container	**27.17 dB**	27.06 dB
Crew	26.73 dB	**27.69 dB**
Football	27.76 dB	**28.15 dB**
Foreman	26.00 dB	**27.06 dB**
Mobile	19.02 dB	**20.19 dB**
News	28.34 dB	**28.60 dB**
Soccer	27.79 dB	**27.98 dB**

Legend



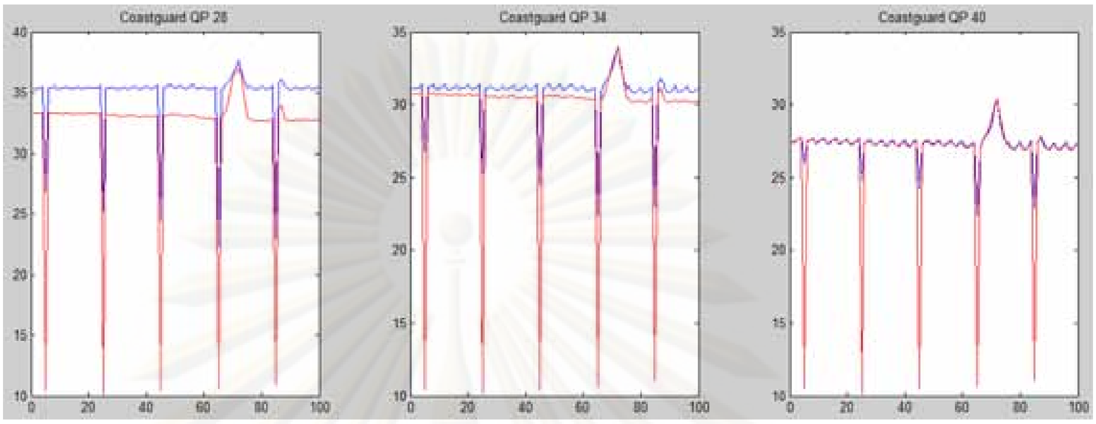


Fig. 4.1 (a). PSNR plot of Coastguard sequence with 5% loss at enhancement layer at various QP.

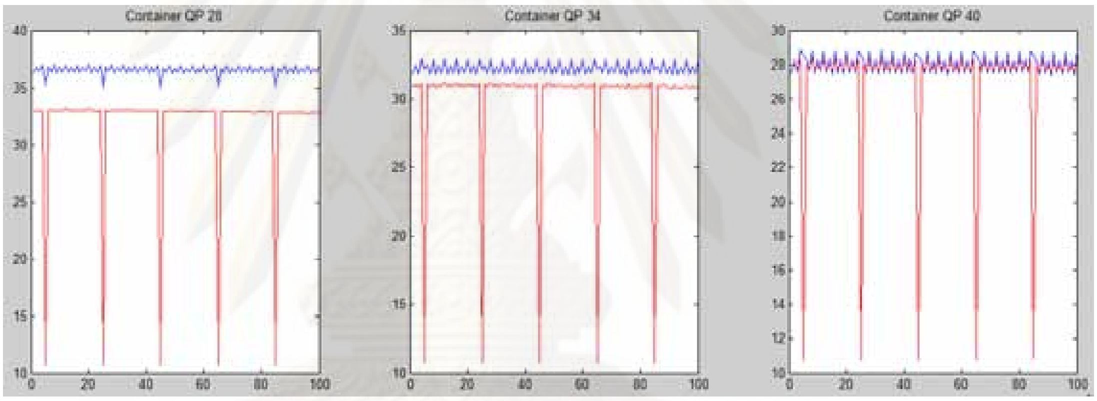


Fig. 4.1 (b). PSNR plot of Container sequence with 5% loss at enhancement layer at various QP.

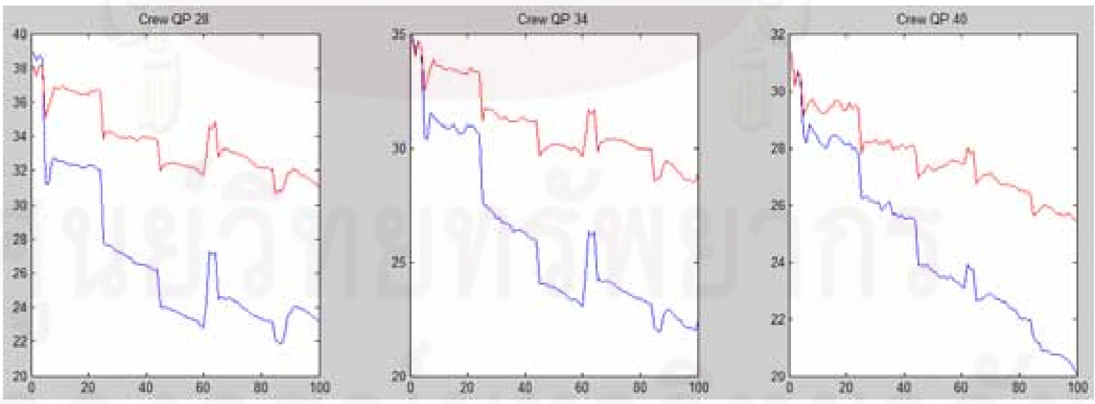


Fig. 4.1 (c). PSNR plot of Crew sequence with 5% loss at enhancement layer at various QP.

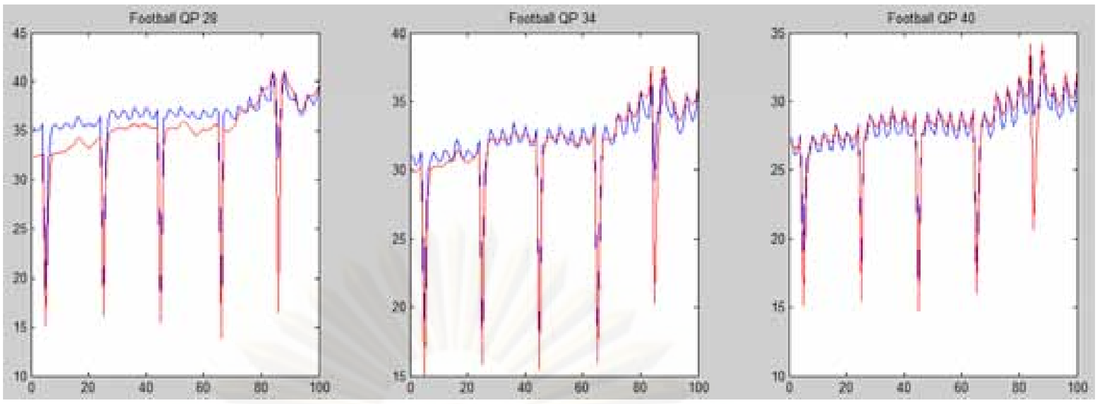


Fig. 4.1 (d). PSNR plot of Football sequence with 5% loss at enhancement layer at various QP.

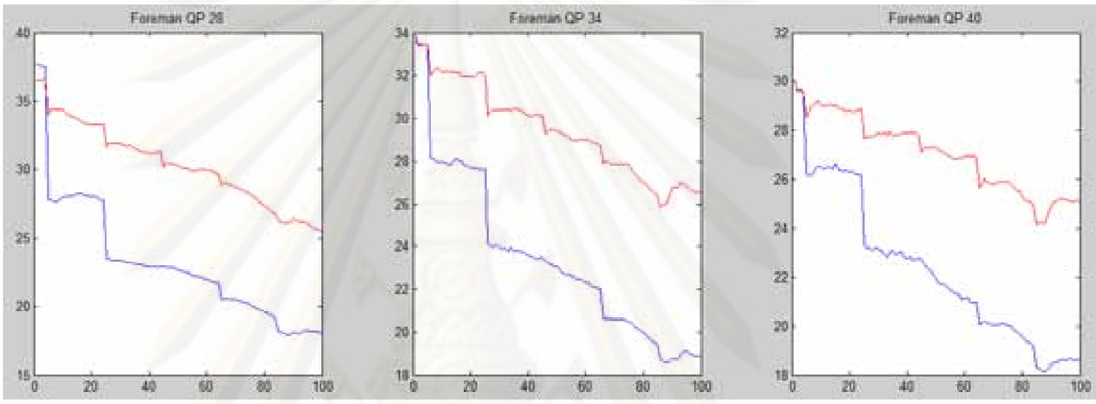


Fig. 4.1 (e). PSNR plot of Foreman sequence with 5% loss at enhancement layer at various QP.

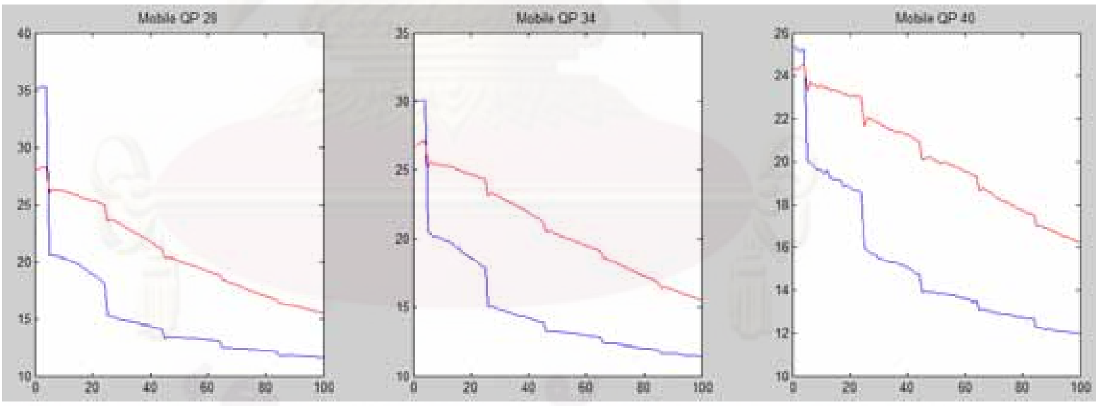


Fig. 4.1 (f). PSNR plot of Mobile sequence with 5% loss at enhancement layer at various QP.

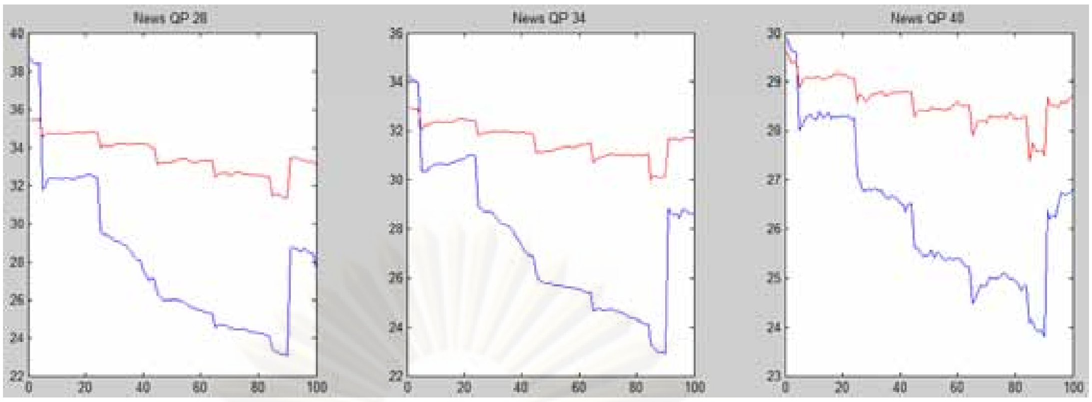


Fig. 4.1 (g). PSNR plot of News sequence with 5% loss at enhancement layer at various QP.

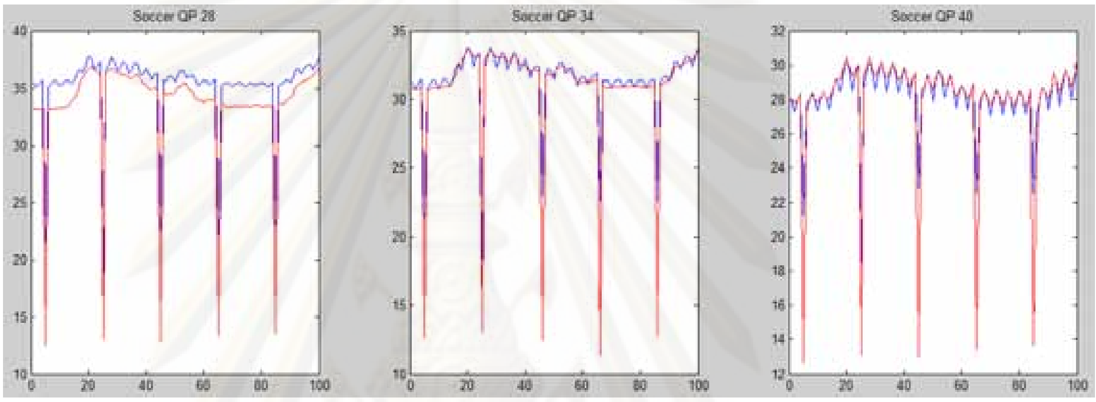


Fig. 4.1 (h). PSNR plot of Soccer sequence with 5% loss at enhancement layer at various QP.

Tables 4.1-3 report the average PSNR after concealment of enhancement layer loss using the hybrid median filter. Significant gains are observed at high motion sequences such as Crew, Foreman and Mobile which are consistently observed at all tested QP values. Also, the proposed algorithm has better performance against BL_Skip in low bit rate video, but the contrary is true for high bit rates.

Figure 4.1 (a-h) gives a more detailed comparison between the BL_Skip method and our proposed algorithm. We are also able to confirm our conclusion with the high motion sequences; as for the other test sequences, we can conclude that our proposed algorithm approaches the performance of the BL_Skip algorithm.

Merely analyzing objective measures of quality such as PSNR cannot fully evaluate the performance of an error concealment algorithm. Subjective evaluation is also important. In the following section, we take sample frames from the concealed sequences in order to assess the visual quality achieved by our concealment algorithm.

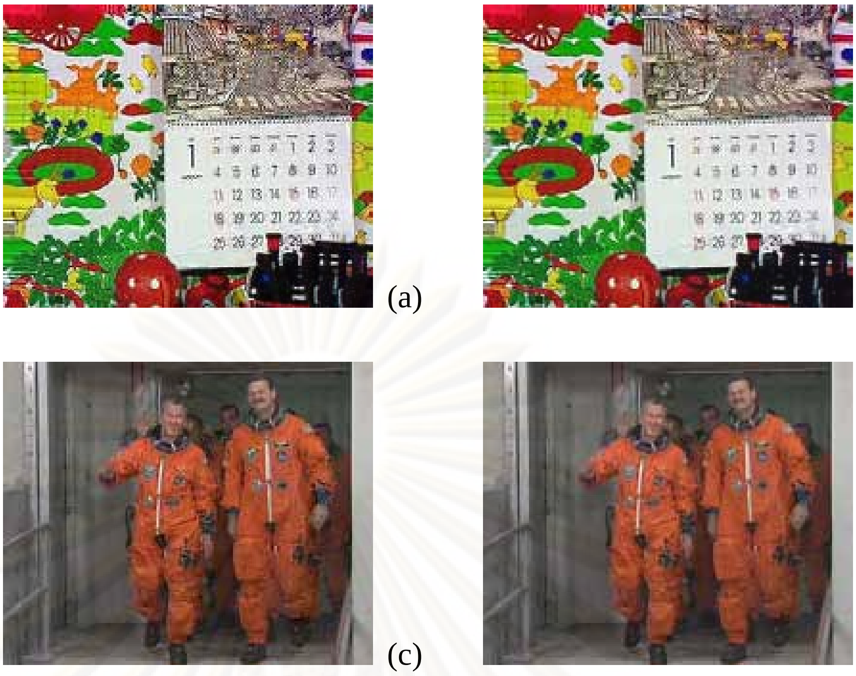


Fig. 4.2. (a) 35th frame of Mobile concealed by BL_Skip. (b) 35th frame of Mobile concealed by proposed method. (c) 48th frame of Crew concealed by BL_Skip. (d) 48th frame of Crew concealed by proposed method. Both sequences originally coded at QP = 28

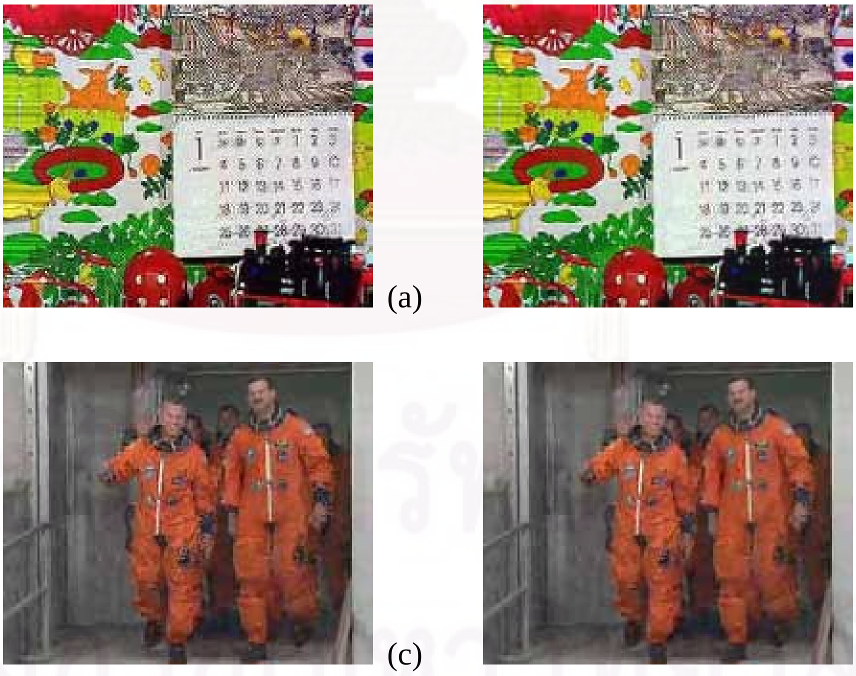


Figure 4.3. (a) 48th frame of Mobile concealed by BL_Skip. (b) 48th frame of Mobile concealed by proposed method. (c) 48th frame of Crew concealed by BL_Skip. (d) 48th frame of Crew concealed by proposed method. Both sequences originally coded at QP = 34

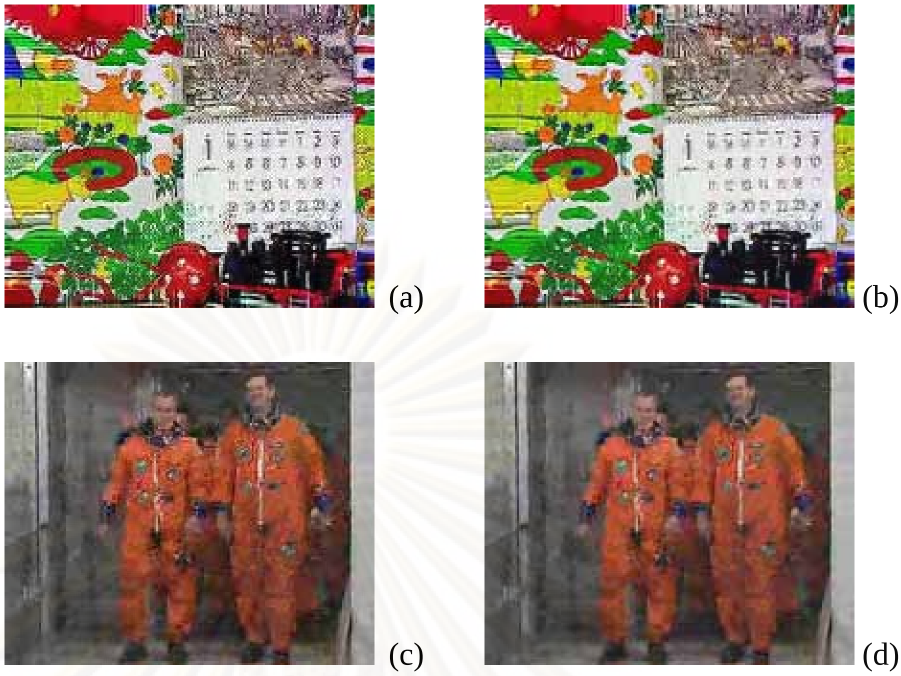


Fig. 4.4. (a) 67th frame of Mobile concealed by BL_Skip. (b) 67th frame of Mobile concealed by proposed method. (c) 54th frame of Crew concealed by BL_Skip. (d) 54th frame of Crew concealed by proposed method. Both sequences originally coded at QP = 40

In the succeeding tests, the encoded bitstream is subject to more severe loss conditions; base layer losses are also considered alongside enhancement layer losses. Both base and enhancement layers are subject to a packet loss rate of 5%. In the event that a base layer frame is lost, any information received in its corresponding enhancement layer frame can be regarded as useless since it depends on the base layer for its proper prediction and decoding. The motion information of the lost base layer frame is estimated from its preceding frame using the weighted motion vector prediction described in the previous section. Only the most recent reference frame is used since it is assumed to have a high correlation with the currently concealed frame. Once the motion vectors are estimated, motion compensated prediction is performed at the enhancement layer. The experimental results are shown below.

Table 4.4. Average PSNR for concealed sequences with 5% loss in base and enhancement layers, QP = 28.

Sequence (QP28)	**BL_Skip**	**Proposed**
Coastguard	**34.95 dB**	32.83 dB
Container	**36.53 dB**	32.61 dB
Crew	26.71 dB	**27.74 dB**
Football	**36.00 dB**	34.97 dB
Foreman	23.11 dB	**24.18 dB**
Mobile	15.27 dB	**16.36 dB**
News	27.98 dB	**28.04 dB**
Soccer	**35.32 dB**	33.97 dB

Table 4.5. Average PSNR for concealed sequences with 5% loss in base and enhancement layers, QP = 34.

Sequence (QP34)	**BL_Skip**	**Proposed**
Coastguard	**30.92 dB**	30.36 dB
Container	**32.25 dB**	30.68 dB
Crew	26.03 dB	**27.06 dB**
Football	31.91 dB	**31.94 dB**
Foreman	23.4 dB	**24.5 dB**
Mobile	14.92 dB	**16.19 dB**
News	27.38 dB	**27.62 dB**
Soccer	**31.5 dB**	31.35 dB

Table 4.6. Average PSNR for concealed sequences with 5% loss in base and enhancement layers, QP = 40

Sequence (QP40)	**BL_Skip**	**Proposed**
Coastguard	**27.28 dB**	27.27 dB
Container	**28.06 dB**	27.77 dB
Crew	24.58 dB	**25.55 dB**
Football	28.03 dB	**28.42 dB**
Foreman	22.3 dB	**23.51 dB**
Mobile	15.18 dB	**16.55 dB**
News	26.33 dB	**26.71 dB**
Soccer	28.23 dB	**28.43 dB**

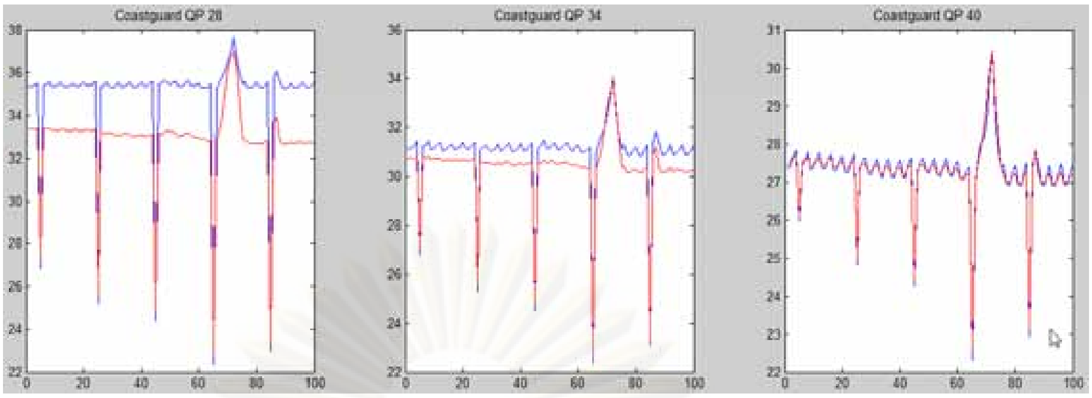


Fig. 4.5 (a). PSNR plot of Coastguard sequence with 5% loss at base and enhancement layer at various QP.

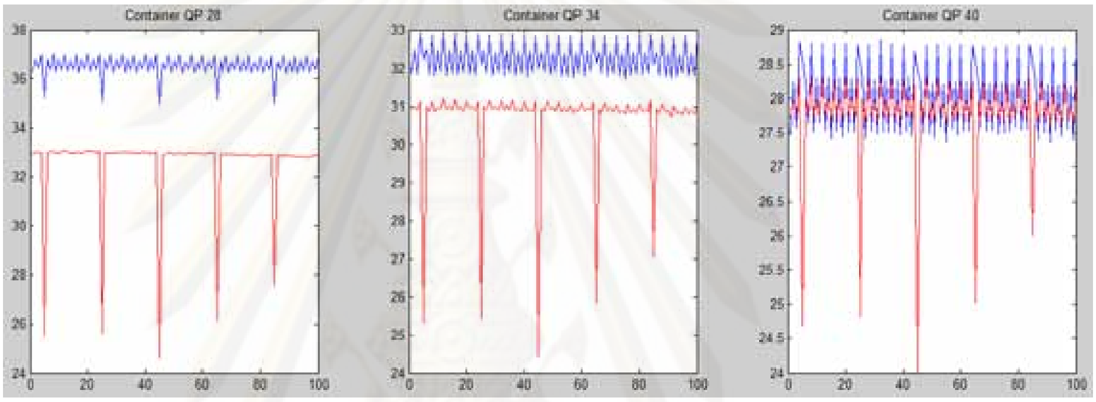


Fig. 4.5 (b). PSNR plot of Container sequence with 5% loss at base and enhancement layer at various QP.

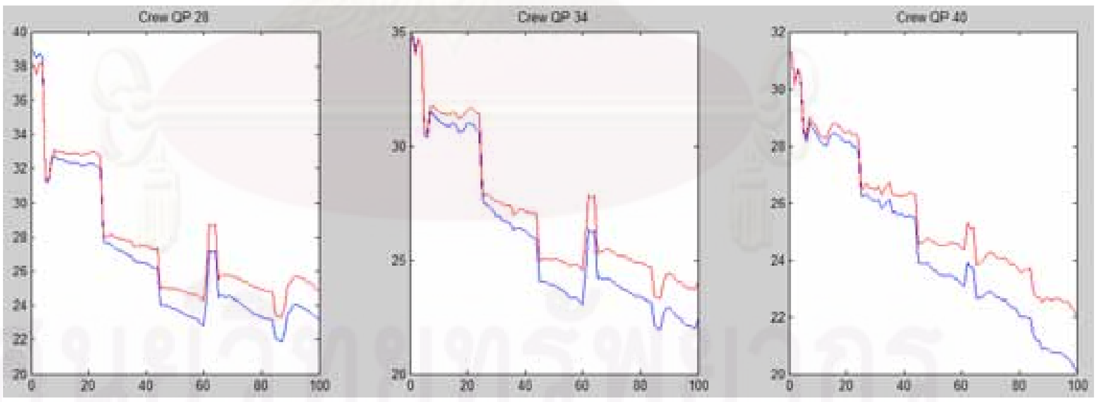


Fig. 4.5 (c). PSNR plot of Crew sequence with 5% loss at base and enhancement layer at various QP.

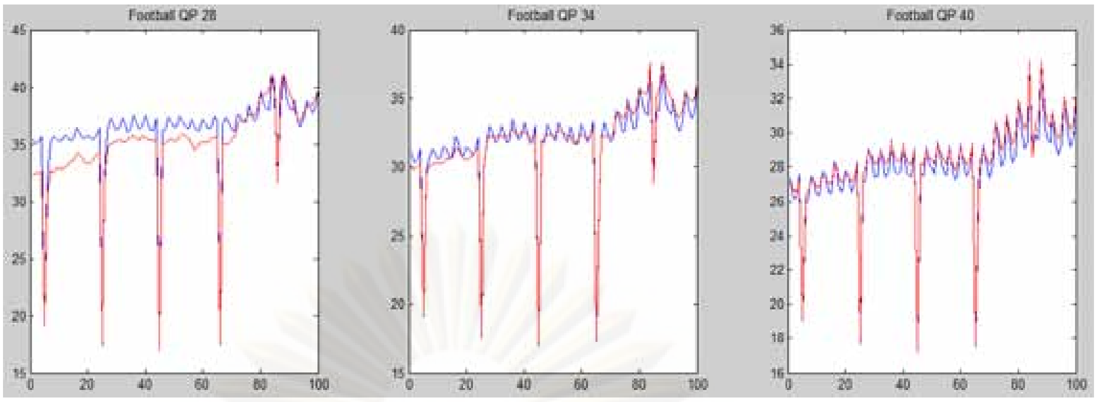


Fig. 4.5 (d). PSNR plot of Football sequence with 5% loss at base and enhancement layer at various QP.

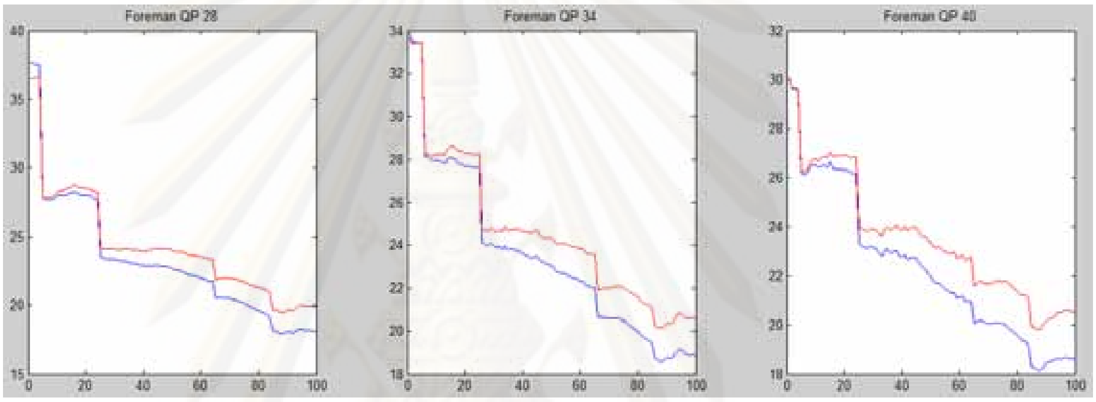


Fig. 4.5 (e). PSNR plot of Foreman sequence with 5% loss at base and enhancement layer at various QP.

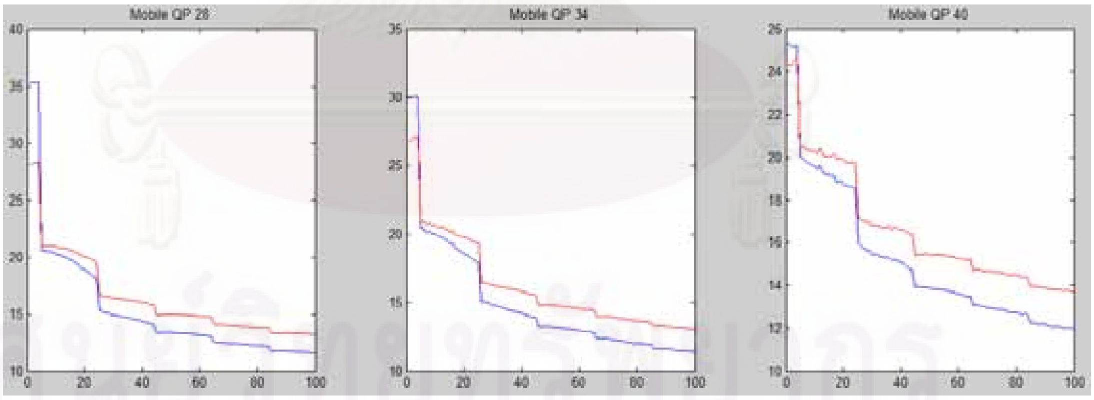


Fig. 4.5 (f). PSNR plot of Mobile sequence with 5% loss at base and enhancement layer at various QP.

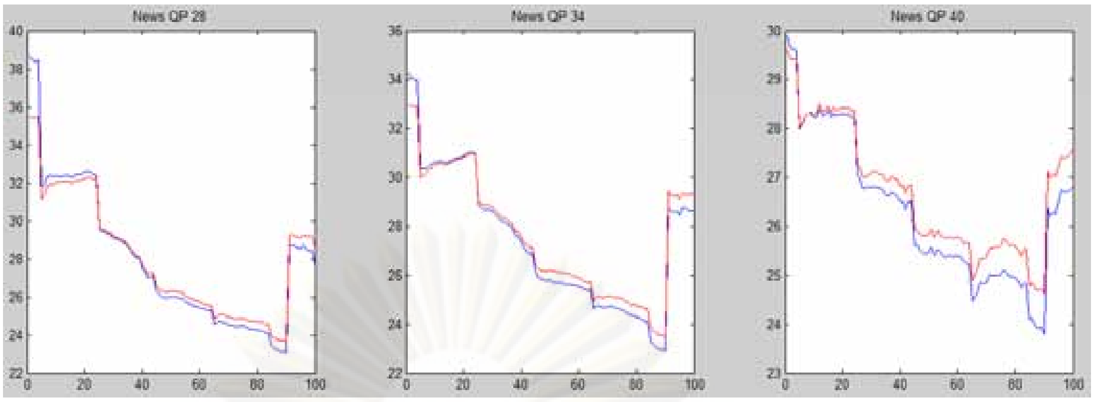


Figure 4.5 (g). PSNR plot of News sequence with 5% loss at base and enhancement layers at various QP.

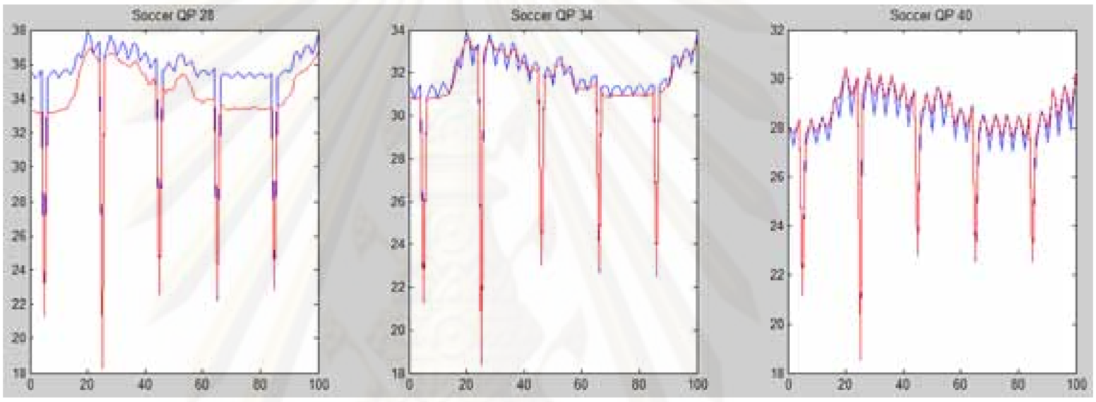


Figure 4.5 (h). PSNR plot of Soccer sequence with 5% loss at base and enhancement layers at various QP.

Evidently, the second test condition is expected to have more severe losses, however, the proposed method still manages to outperform the BL_Skip algorithm in most of the test scenarios although the margin is smaller. Once again, we evaluate the subjective output of the proposed algorithm.

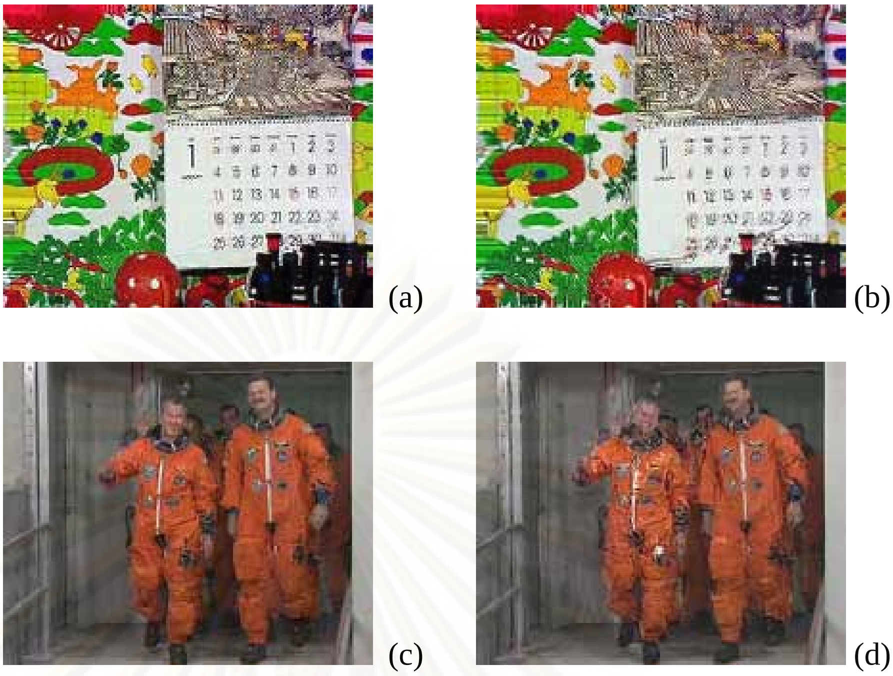


Fig. 4.6 (a) 35th frame of Mobile concealed by BL_Skip. (b) 35th frame of Mobile concealed by proposed method. (c). 48th frame of Crew concealed by BL_Skip. (d) 48th frame of Crew concealed by proposed method. Both sequences originally coded at QP = 28

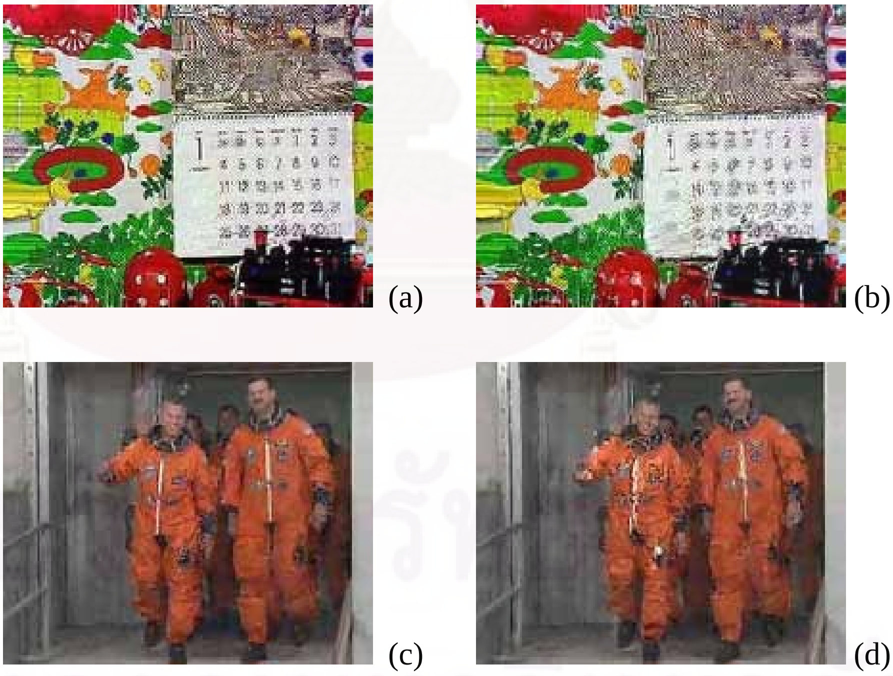


Fig. 4.7. (a) 48th frame of Mobile concealed by BL_Skip. (b) 48th frame of Mobile concealed by proposed method. (c) 48th frame of Crew concealed by BL_Skip. (d) 48th frame of Crew sequence concealed by proposed method. Both sequences originally coded at QP = 34

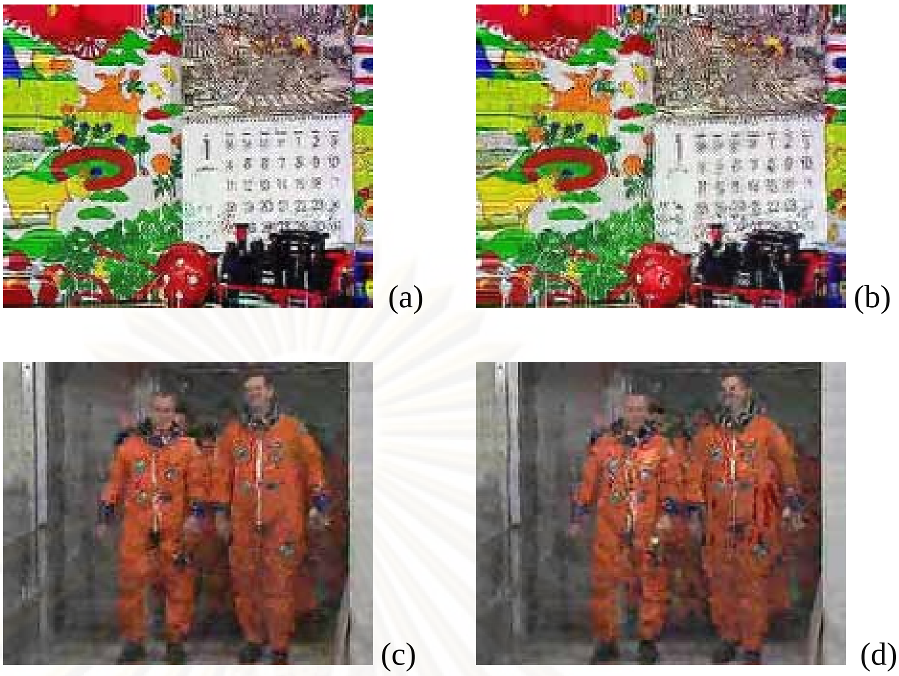


Fig. 4.8. (a) 67th frame of Mobile concealed by BL_Skip. (b) 67th frame of Mobile concealed by proposed method. (c) 54th frame of Crew concealed by BL_Skip. (d) 54th frame of Crew concealed by proposed method. Both sequences originally coded at QP = 40

CHAPTER V

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

Based on the PSNR metrics, our proposed algorithm generally outperformed the conventional BL_Skip algorithm. In the first case where the base layer is intact, we chose to simply upsample the base layer motion information since it corresponds to the lost enhancement layer frame. This way, we are also able to reduce the computations for motion estimation. Applying the hybrid median filter on the reconstructed frame was found to be effective for removing artifacts left by the concealment process. An average gain of about 0.6-0.8 dB especially for the high bit rate transmission is achieved. The most significant gains are seen in the sequences Crew, Foreman and Mobile. These clips all have fast and complex motion, which is one of the weak points of the BL_Skip algorithm. The PSNR curve also reveals that error propagation is also reduced. Subjective tests reveal that the proposed method is better at preserving the edges and shapes of objects in the scene. Artifacts are effectively suppressed hence giving a clearer picture to the viewer. Some loss of texture might be observed due to the smoothing effects of the filter. The proposed algorithm was found to be more effective in higher QP since texture is no longer as significant at low bit rates.

In the second set of tests, we are able to test the performance of the proposed motion estimation scheme against the BL_Skip. Both the base layer and enhancement layer frames are lost hence motion must be estimated from the temporal neighbors. We considered the most recent reference frame for our motion estimation scheme since the original sequence was also encoded using a single reference frame. This suggests that it would also have the highest correlation in terms of motion and texture content to the lost frame. Looking at the results, we can still see that the proposed algorithm has the edge over BL_Skip. The gain is not as great since the loss scenario in this test is worse. As in the first test, we observe that the proposed method well outperforms the BL_Skip algorithm in high motion sequences such as Crew, Foreman and Mobile. Subjective tests show that there is some smudging of texture resulting from the median filtering operation. The BL_Skip algorithm is able to retain the details but still manifests many artifacts. Overall, the proposed algorithm has an average gain of 0.2-0.6 dB especially for the low bit rate case.

We can clearly say that the proposed algorithm generally performs better than BL_Skip algorithm. The motion estimation scheme proved to perform well in the high and complex motion sequences. Also, judging by the PSNR curves, we can see that the proposed algorithm can reduce the effect of error propagation.

5.2 Future Works

In this paper, we have proposed an error concealment scheme using motion estimation and post processing. We have successfully outperformed the conventional BL_Skip algorithm. However, to further improve the concealment method, it is possible to refine the motion estimation scheme by using a smaller block size. A smoother motion estimate may also be gained by using multiple reference frames; this may help reduce blocking artifacts. Another way of avoiding blocking artifacts is to use a side matching distortion measure when considering the motion vector estimate. Aggressive median filtering was seen to have caused loss in texture and detail in the reconstructed frame. It may be possible to selectively apply the median filter only on noise affected areas of the image. The process may require segmentation techniques and are outside the scope of our study. In line with recovery of detail and texture, estimation of the residual can also be studied.

In the light of multimedia content distribution, it is desirable to integrate the decoder error concealment with error resilience tools at the encoder. Error resilience tools can aid the performance of error concealment by preventing severe information loss. If more information is available at the decoder, there are more prediction sources which can be used by the error concealment method.

References

[1] Baccichet, P., Bagni, D. Chimienti, A., Pezzoni, L., & Rovati, F.S. (February 2005). Frame concealment for H.264/AVC decoders. *IEEE Transactions on Consumer Electronics*. 51 (1), 227-233.

[2] Baccichet, P., & Chimienti, A. (2004). A low complexity concealment algorithm for the whole frame loss in H.264/AVC. *2004 IEEE 6th Workshop on Multimedia Signal Processing.*

[3] Belfiore S., Grangetto, M., Magli, E., & Olmo, G., (April 2005). Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation. *IEEE Transactions on Multimedia.* 7 (2), 316-329.

[4] Chen Y., Yu, K., Li, J., & Li, S., (2004). An error concealment algorithm for entire frame loss in video transmission. *Microsoft Research Asia.*

[5] Chen Y., Xie, K., Zhang, F., Pandit, P., & Boyce, J., (2006). Frame loss error concealment for SVC. *Journal of Zhejiang University SCIENCE A.* 7 (5). 677-683.

[6] Fan, J., & Zhang, X., (2007). A new error concealment scheme for whole frame loss in video transmission. *Picture Coding Symposium*.

[7] Garcia D., (November 2009). Hybrid Median Filtering. http://www.biomecardio.com/matlab/hmf.html. [March 16, 2010].

[8] Ghandi, M.M., & Ghandbari, M., (February 2006). Error concealment for SNR scalable video coding. *Signal Processing: Image Communication.* 21 (2). 91-99.

[9] Hui, L., Wang, Y., & Li., H. (August 2008). A comparison between SVC and transcoding. *IEEE Transactions on Consumer Electronics.* 54 (3). 1439-1446.

[10] Kim, J., Hwang, M., Yang, H., Ko, S., Morales, A.W., (January 2008). Frame error concealment technique using adaptive inter mode estimation for H.264/AVC. *ICCE 2008.*

[11] Liu, Y., Bu, J., Chen, C., Mo, L., & He, K., (2007). Multiframe error concealment for whole-frame loss in H.264/AVC. *ICIP 2007.* 4. IV-281-IV284.

[12] Ma, Q. Wu, F. Lou, J., & Sun, M., (May 2009). Frame loss error concealment for spatial scalability using hallucination. *17th International Packet Video Workshop 2009.*

[13] Quacchio, E., Magli, E., Olmo, G., Baccichet, P., & Chimienti, A. (March 2005). Enhancing whole frame error concealment with an intra motion vector estimator in H.264/AVC. *International Conference on Acoustics, Speech, and Signal Processing.* 329-332.

[14] Rimac-Drlje, S., Nemcic, O., & Vranjes, M., (September 2008). Scalable video coding extension of the H.264/AVC standard. *ELMAR 2008.*

[15] Schwarz, H., Marpe, D., & Wiegand, T., (September 2007). Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology.* 17 (9). 1103-1120.

[16] Schwarz, H., & Wien, M., (March 2008). The scalable video coding extension of the H.264/AVC standard: standards in a nutshell. *IEEE Signal Processing Magazine.* 5. 135-141.

[17] Song, K., Chung, T., Oh, Y. Kim, C., & Kim, Y., (May 2007). Error concealment of H.264/AVC video frames for mobile video broadcasting. *IEEE Transactions on Consumer Electronics.* 53 (2). 704-711.

[18] Sui, S., Liu, F., & Xue, P., (August 2007). Whole frame error concealment with improved backward motion estimation for H.264 decoders. *ICIG 2007.*

[19] Tang, T., Wang, J., Liu, Y., Gao, Y., & Yu, S., (October 2007). Adaptive frame recovery based on motion activity. *IEEE Workshop in Signal Processing Systems.*

[20] Xu, Y., & Zhou, Y., (2004). Refined video error concealment over wireless IP network. *Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication.*

[21] Yi, G., Chen, Y., Wang, Y. Li, H. Hannauksela, M.M., & Gabbouj, M., (June 2009). Error resilient coding and error concealment in scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology.* 19 (6). 781-795.

[22] Lam, S & Aramvith, S., (December 2009). Error concealment of whole frame loss in H.264/AVC using improved motion vector estimation. *International Conference on Science and Engineering*, 1 (27), 127-129.

[23] Lam, S, Cajote, R., & Aramvith, S.. (March 2010). Error concealment in H.364 spatial scalable video coding using improved motion vector estimation. *International Conference on Green Computing.* I (1). 140-144.

APPENDICES

APPENDIX A: PUBLICATION IN ICSE 2009

Error Concealment of Whole Frame Loss in H.264/AVC Using Improved Motion Vector Estimation

Simon Jude Q. Lam[1], and Asst. Prof. Supavadee Aramvith[2]
Department of Electrical Engineering, Chulalongkorn University, 54 Phayathai Road, Patumwan, Bangkok, Thailand. 10330
[1]simonjudelam@gmail.com
[2]supavadee.a@chula.ac.th

***Abstract*— H.264/AVC provides the capability to encode highly compressed videos by exploiting spatiotemporal redundancy. High compression is needed to transmit videos over bandwidth restricted channels, but also increases its vulnerability to transmission errors. The loss of one packet may lead to the loss of an entire frame, and subsequently degrading the quality of the following frames. We propose an error concealment algorithm for H.264/AVC that will estimate the lost frame based on motion information from the previous frames. We give importance to the frames close to the lost frame, but also consider the motion of other previous frames. The results show some improvement over conventional methods of concealing frame losses such as frame copy and motion copy methods.**

***Keywords*— H.264, Error Concealment, Motion Vector Estimation, Frame Loss, Transmission Error**

I. INTRODUCTION

Video transmission over lossy networks, such as a wireless network, may degrade the visual quality of the decoded video. Of special interest is the highly compressed video using the latest video coding standard, H.264/AVC. H.264/AVC video can suffer from entire frame loss in the event of a transmission error. This is because an entire video frame can be encoded into a single network packet [3]. In addition, the succeeding frames also become affected by this error due to the spatiotemporal dependencies among neighboring frames in H.264/AVC [5].

Error concealment at the decoder can be used in order to reduce the effects of such errors. It requires no additional information nor modification at the encoder side [6]. Although the recent versions of JM reference software internally support error concealment algorithms, their performance are often considered inadequate.

In the scenario of a frame loss, one cannot simply apply simple error concealment techniques. Let us assume that frame *n* of a video sequence was lost during transmission. It is not possible to perform spatial error concealment on the macroblocks of frame n, since it requires the data from neighboring macroblocks on the same frame. This is due to the fact that the whole frame is lost. One of the common techniques in temporal error concealment is the Frame Copy (FC) method. In this technique, the pixels of the frame *n-1* are duplicated onto the lost frame. The motion vectors for this frame are set to zero. This technique may work for stationary regions of the frame, but fails on the opposite scenario [4]. Another technique is the Motion Copy (MC) method. In this case, the motion vectors for each macroblock of the previous frame are projected onto the lost frame. The lost frame is estimated using these motion vectors and an assigned reference frame. However, MC method often results to blocking artifacts in the concealed frame, due to the block-based prediction.

In [2], Belfore, et al. proposed an error concealment scheme for MPEG-2 that utilizes the motion information of up to *L* past frames. The underlying assumption of their work characterizes the motion of each pixel following a linear trajectory. For each pixel in frame *n-1*, a motion vector history is generated. Starting from frame *n-1*, the motion vectors of each pixel between frame *n-1* and frame *n-2* are recorded. The process is repeated until the *L*th frame or until the motion vectors point to a valid location. Afterwards, a motion vector estimate is formed in order to predict frame *n* from frame *n-1*. The predicted motion vector for each pixel in frame *n-1* is taken as the average motion vector of its motion vector history. Any pixels on frame *n* with no projected motion vector will be filled by applying a median filter around its neighborhood. The linear motion assumption has proven to be sufficient for their purposes [2]. In [1], the motion vectors are estimated in a similar manner with [2], but the estimation is performed at the block level in order to reduce computational complexity.

Several researches have used [1] and [2] as a benchmarks for their error concealment schemes, and are commonly referred to as CA_B (concealment algorithm on blocks) and CA_P (concealment algorithm on pixels) respectively. For example, in [8], recognizes the effective performance of CA_P but also points out that the method proposed by Belfiore, et al. is computationally expensive, mainly because of repeated median filtering operations. One of the proposed ideas in [8] also estimated the motion vectors at the pixel level, however, it is done by taking a weighted average of the motion vectors of the macroblocks in a neighborhood. With CA_B, the number of computations is reduced since motion vectors are estimated on the block level.

While the above algorithms mainly consider the motion vectors on the same position in the previous frames, other algorithms estimate the motion vectors whose projection falls on the currently concealed macroblock. Such an example can be seen in [9]. Motion vectors are estimated from the temporal neighbors of the frame to be concealed. For instance, the macroblocks of frame *n-1* are projected onto the lost frame *n*; it is possible that a macroblock in frame *n* may have multiple macroblock projections from frame *n-1*. The motion vectors corresponding to these macroblocks will be used to generate an estimated motion vector. The forward motion vector estimate is calculated as a weighted average of the pertinent motion vectors. The weights correspond to the area in pixels covered by the macroblock's projection. [10] shows how the coverage of the projection of the extrapolated motion vectors influence a damaged area of a corrupted frame.

In this paper, we propose an error concealment scheme that uses the motion information in multiple previous frames. However, unlike in [2], we also consider the influence of the neighboring macroblocks in our motion estimation method, since there is spatial correlation among neighboring blocks. Motion vectors from the collocated macroblock, as well as the macroblock on top and to the left are assigned weights relative to their displacements in the horizontal and vertical directions, instead of the overlapped area.

The rest of the paper is organized as follows: Section II describes the proposed motion estimation scheme, while Section III presents the experiments and results. Finally, a conclusion is found in Section IV.

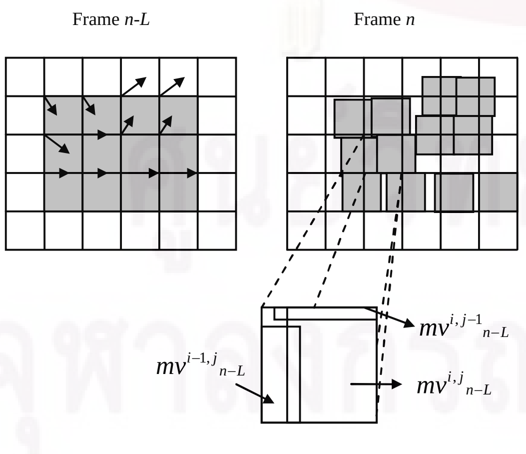
II. PROPOSED BLOCK-BASED WEIGHTED MOTION ESTIMATION

In H.264/AVC, it is possible to use multiple reference frames in video encoding. For each macroblock, a best match can be searched from frames other than the previous frame. We want to exploit this feature in order to generate a more accurate estimate of the motion vectors. For our experiments, we have set the number of reference frames *L* equal to 5. Block-based approach is also used to lessen the computational cost. We tried to compare the performance of simply taking the average motion vector of the past *L* frames and assigning different weights to each frame. Our assumption is that the correlation of motion information is higher for adjacent frames than of frames with further temporal distance. We assign the weights of the motion vectors with linear decay as shown in eq. (1). Our experiments show that assigning different weights is better than assigning equal weights to the motion vectors.

$$mv^{i,j}{}_n = \frac{1}{15}\begin{pmatrix} 5\times(m\hat{v}^{i,j}{}_{n-1}) + 4\times(m\hat{v}^{i,j}{}_{n-2}) + \\ 3\times(m\hat{v}^{i,j}{}_{n-3}) + 2\times(m\hat{v}^{i,j}{}_{n-4}) + \\ (m\hat{v}^{i,j}{}_{n-5}) \end{pmatrix} \quad (1)$$

Where $mv^{i,j}{}_n$ is the motion vector used for predicting the macroblock at (i, j) in the lost frame. $m\hat{v}^{i,j}{}_{n-L}$ is calculated as a weighted average of $mv^{i,j}{}_{n-L}$, $mv^{i-1,j}{}_{n-L}$, *and* $mv^{i-1,j}{}_{n-L}$.

Let $m\hat{v}_x{}^{i,j}{}_{n-L}$ and $m\hat{v}_y{}^{i,j}{}_{n-L}$ be the horizontal and vertical components of $m\hat{v}^{i,j}{}_{n-L}$. $m\hat{v}_x{}^{i,j}{}_{n-L}$ and $m\hat{v}_y{}^{i,j}{}_{n-L}$ in the *L*th previous frame is computed as shown in eqs. (2) and (3):



$$\begin{aligned} m\hat{v}_x{}^{i,j}{}_{n-L} &= \alpha_{LEFT} mv_x{}^{i-1,j}{}_{n-L} + \alpha_{TOP} mv_x{}^{i,j-1}{}_{n-L} \\ &+ \alpha_{CENTER} mv_x{}^{i,j}{}_{n-L} \end{aligned} \quad (2)$$

$$m\hat{v}_y{}^{i,j}{}_{n-L} = \alpha_{LEFT} mv_y{}^{i-1,j}{}_{n-L} + \alpha_{TOP} mv_y{}^{i,j-1}{}_{n-L} + \alpha_{CENTER} mv_y{}^{i,j}{}_{n-L} \quad (3)$$

Where α is a weighting factor computed as follows:

$$\alpha_{:LEFT} = \frac{mv_c{}^{i-1,j}{}_{n-L}}{mv_c{}^{i-1,j}{}_{n-L} + mv_c{}^{i,j-1}{}_{n-L} + mv_c{}^{i,j}{}_{n-L}} \quad (4)$$

$$\alpha_{:TOP} = \frac{mv_c{}^{i,j-1}{}_{n-L}}{mv_c{}^{i-1,j}{}_{n-L} + mv_c{}^{i,j-1}{}_{n-L} + mv_c{}^{i,j}{}_{n-L}} \quad (5)$$

$$\alpha_{:CENTER} = \frac{MB\_size - mv_c{}^{i,j}{}_{n-L}}{mv_c{}^{i-1,j}{}_{n-L} + mv_c{}^{i,j-1}{}_{n-L} + mv_c{}^{i,j}{}_{n-L}} \quad (6)$$

In eqs. (4), (5), and (6), the subscript *c* indicates for which component (x or y) the weight is computed for. Hence, the horizontal and vertical components are weighed differently. MB_size indicates the macroblock size which is kept to 16x16 in this paper. We consider the motion of the top, left, and collocated macroblocks wince the spatial correlation among them is high. The estimated motion vectors are used for motion compensated prediction of frame *n* from frame *n-1*.

III. EXPERIMENTS AND RESULTS

In order to evaluate the performance of the proposed algorithm, we ran a test on the following conditions. The encoded videos are in QCIF resolution at 30 fps. The number of reference frames is set to 5, and using only 16x16 blocks for motion compensated prediction. Finally, we encapsulate each video frame into one slice group and set the bitrate at 128000 bps. These videos were then subject to a packet loss rate of 5%. The video sequences used for this test are Akiyo, Carphone, Coastguard, Foreman, and Mobile.

The proposed concealment algorithm is implemented in JM 14.2. We compare the performance of our proposed algorithm with the motion copy method.

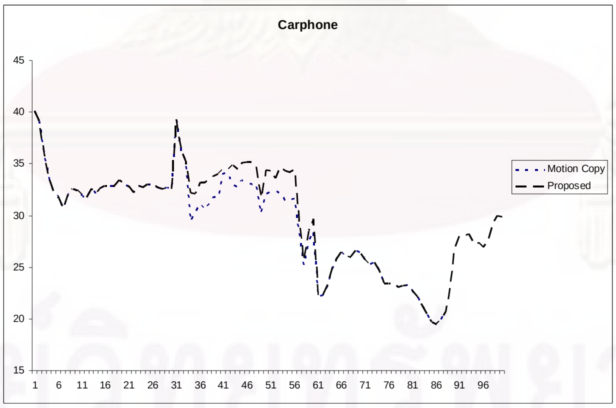


Figure 1. PSNR plot of Carphone sequence.

Fig. 2 Frame 33 of Carphone sequence. (a) No Error, (b) Motion Copy, (c). proposed method.

As shown in the results, there is generally an improvement in the PSNR. Particularly in the Carphone sequence, the proposed algorithm was able to realize an improvement of around 2 dB for the lost frame. In addition, the succeeding frames also exhibited around 1dB improvement, showing our method's potential to suppress error propagation.

Fig. 2 shows a comparison among the error-free frame (a), the lost frame concealed by motion copy (b), and the same frame concealed by our proposed method (c). In the Motion Copy method, we can see some blocking artifacts typically arising from conventional block-based concealment methods. IN our method, the blocking artifacts are less visible since we consider the spatial correlation of neighboring macroblocks.

TABLE I
AVERAGE PSNR FOR CONCEALED SEQUENCES

Sequence	**Motion Copy**	**Proposed**
AKIYO	37.06 dB	**37.20 dB**
CARPHONE	29.21 dB	**29.74 dB**
MOBILE	21.96 dB	**22.00 dB**

IV. CONCLUSION

In this paper, we have seen the potential of improving the performance of error concealment by reformulating the manner in which motion is estimated from a correctly received frame to a lost frame. We have found that by considering the temporal distance, a better motion estimate can be formulated. In addition, we found that the motion information found in the macroblock neighborhood can be used to further improve the estimated motion vector for the lost frame. Furthermore, the same technique may find potential application in the scalable extension of H.264. H.264/SVC is also capable of performing motion vector refinement at quarter-pel resolution which may improve the accuracy of the estimated motion vectors, potentially for intra layer and inter layer prediction modes.

REFERENCES

[1] P. Baccichet, D. Bagni, A. Chimienti, L. Pezzoni, and S. Rovati, "*Frame concealment for H.264/AVC decoders*," IEEE Trans. Consumer Electron., vol. 51, no. 1, pp. 227-233, Feb. 2005.

[2] S. Belfore, M. Grangetto, E.Magli, G. Olmo, "*An error concealment algorithm for streaming video*," in Proc. ICIP 2003, 2003

[3] Y. Liu, J. Bu, C. Chen, L. Mo, K. He, "*Multiframe Error Concealment for Whole-Frame Loss in H.264/AVC.*" ICIP (4) 2007: 281-284

[4] T. Tang, J. Wang, Y. Liu, Y. Gao, S. Yu, "*Signal Processing Systems*", 2007 IEEE Workshop, Oct. 2007 Page(s):692 – 697

[5] Y. Wang, J. Ostermann, and Y.Q. Zhang, *Video Processing and Communication.*, Signal Processing Series, NJ: Prentice Hall, 2002.

[6] Y. Wang, S. Wenger, J. Wen and A. Katsaggelos, "*Error Resilient Video Coding Techniques,*" IEEE Signal Processing Magazine, pp. 61-82, July 2000.

[7] P. Baccichet, A. Chimienti, *"A low complexity concealment algorithm for the whole-frame loss in H.264/AVC,"* Multimedia Signal Processing, 2004 IEEE 6th Workshop on , vol., no., pp. 279-282, 29 Sept.-1 Oct. 2004

[8] K. Song, T. Chung, Y. Kim, Y. Oh, Chang-Su Kim, "*Error Concealment of H.264/AVC Video Frames for Mobile Video Broadcasting,*" Consumer Electronics, IEEE Transactions on , vol.53, no.2, pp.704-711, May 2007

[9] J.H. Kim, M.C. Hwang, H.Y. Yang; S.J. Ko; Morales, A.W., *"Frame Error Concealment Technique Using Adaptive Inter-Mode Estimation for H.264/AVC,"* Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on , vol., no., pp.1-2, 9-13 Jan. 2008

[10] Q. Peng, T. Yang, C. Zhu, *"Block-based temporal error concealment for video packet using motion vector extrapolation,"* Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on , vol.1, no., pp. 10-14 vol.1, 29 June-1 July 2002

APPENDIX B: PUBLICATION IN ICGCS 2010

Error Concealment in H.264 Spatial Scalable Video using Improved Motion Estimation

Simon Jude Q. Lam *), Supavadee Aramvith **)
*) Department of Electrical Engineering
Chulalongkorn University, 54 Phayathai Road,
Patumwan, Bangkok, Thailand, 10330
simonjudelam@gmail.com
**) Department of Electrical Engineering
Chulalongkorn University, 54 Phayathai Road,
Patumwan, Bangkok, Thailand, 10330
supavadee.a@chula.ac.th

Abstract—**The scalable video coding extension of the H.264/AVC standard includes support for spatial scalability. When a scalable video bit stream is transmitted over a packet network, it may be subject to frame losses at high compression ratios, thus degrading the quality of decoded video. Information from temporally neighboring frames may be used to estimate the motion information in a lost frame. In this paper, we test our proposed motion vector estimation against conventional error concealment methods applied in H.264/AVC. Results show that our method has a maximum gain of 4 dB over the conventional Frame Copy and is reflected in subjective evaluation of the concealed video sequences.**

Keywords-error concealment; motion vector estimation; H.264/AVC scalable video coding extension

INTRODUCTION

Spatial scalability provides a means for transmission of video targeted at a variety of decoding terminals. With great improvements in coding efficiency, the scalable extension of H.264/AVC provides a more viable solution compared to its predecessors [1]. However, the scalable video bitstream is still prone to losses when transmitted over a packet network. In the worst case scenario, an entire video frame may be lost since the H.264 codec is capable of compressing one frame into one packet. In addition, the use of predictive coding among consecutive video frames increases the effects of error propagation [2]. Implementing error resilience at the encoder cannot guarantee recovery from all errors. Hence, it is still desirable to implement error concealment at the decoder. Our proposed method aims to conceal the loss of entire frames.

PREVIOUS WORK ON ERROR CONCEALMENT

Non-normative Error Concealment

In 2005, Chen et al. [3] proposed a set of non-normative error concealment schemes for the scalable video coding extension of H.264/AVC which were later adopted into the JSVM reference software.

First of their methods is the Frame Copy (FC) method. FC can be considered as an intra-layer concealment method since it does not require information from dependent layers. FC copies the pixel information from the most recent frame stored in reference List_0. It is considered to be the least complex but also least effective especially in motion filled sequences.

Secondly, there is the Temporal Direct (TD) method. The macroblocks to be concealed are assumed to be coded in the temporal direct mode. Similarly, TD also does not depend on information in other layers in order to perform concealment. TD was found to work best in sequences with smooth and simple motion.

The final method implemented in the JSVM is the motion and residual upsampling method which is simply called BL_Skip. As its name implies, it uses an upsampled version of motion information from a lower layer to perform motion compensated prediction at the upper layer. The residual information of the lower layer is also upsampled using a 4-tap filter for use in the upper layer following motion compensated prediction. However, when the base layer information is not available, the base layer frame is first concealed using the TD method prior to upsampling its motion information. Hence, BL_Skip is only as good as TD when concealing base layer losses. In addition, BL_Skip potentially yields blurred videos due to the use of upsampled information from the base layer residual.

Frame Error Concealment Technique using Adaptive Inter Mode Estimation

Hwang M.C. et al. presented a robust error concealment technique for whole frame losses in H.264/AVC [4]. This technique is especially useful when spatial error concealment is not possible due to losses of huge regions of a frame. Hwang initially designed a bidirectional approach to the motion estimation problem. For a lost macroblock, forward projections from the previous frame and backward projections from the future frame are checked. The motion vectors whose projections fall on the area of the macroblock to be concealed are considered. It can be expected that multiple projections can fall onto a single macroblock, hence producing multiple motion vector candidates. The final motion vector estimate is formulated as a weighted average of all the motion vector candidates. Weights are assigned to the motion vector candidates corresponding to the area of the overlapping region. Motion estimation is carried out on 4x4 sized blocks. In some cases, one or more 4x4 blocks may have no motion vector projection. In Hwang's paper, it is assumed that each macroblock is encoded in one of the H.264/AVC inter modes; as such the inter mode of the concealed macroblock is also estimated adaptively in order to assign motion vectors to all 4x4 blocks.

In our paper, we wish to improve on the rather simple error concealment schemes in JSVM by applying an improved motion vector estimation scheme.

PROPOSED MOTION ESTIMATION SCHEME

The accuracy of motion vectors greatly influences the performance of the error concealment scheme. We wish to exploit the high correlation on motion information between successive video frames. Specifically, we will use the motion information from the previous frame in order to estimate the lost motion vector information in the currently concealed frame.

In the Fig. 1 below, let the center macroblock be a lost macroblock which needs to be concealed in the current frame. Macroblocks A to I represent the projections of the macroblocks in the previous frame onto the current frame. Macroblocks B, D, and E clearly overlap with the macroblock to be concealed; the overlapped area is marked as the dark gray area. Therefore the motion vectors of B, D, and E will be considered for estimating the motion vector of the lost macroblock.

Figure 1. Macroblock projections from previous frame.

For the currently concealed macroblock, the motion vector estimate is formulated as follows:

$$\widetilde{MV} = \frac{\sum_{i=0}^{N} MV_i \cdot W_i}{\sum_{i=0}^{N} W_i}. \quad (1)$$

Where N is the number of macroblocks with overlapping areas, and W is the weighting factor assigned to the *i*th motion vector belonging to the *i*th macroblock. *W* is counted as the area in pixels which the *i*th macroblock overlaps with the currently concealed macroblock.

TEST AND RESULTS

The proposed motion estimation scheme is implemented in the JSVM reference software version 8.5. Test sequences are encoded with spatial scalability; the base layer is encoded at QCIF resolution and the enhancement layer is encoded at CIF resolution. The coding sequence is IPPP.

We investigate the effectiveness of our proposed algorithm against the conventional Frame Copy method. The encoded video sequences are subject to 10% packet loss rate which is effected on both base and enhancement layers.

Fig. 2-4 plots the PSNR for the first 100 frames of Crew, Foreman, and News respectively. It is visible that there is a significant improvement over the frame copy method. In majority of the tested video sequences, the proposed method outperforms the conventional Frame Copy method with a maximum average gain of 4dB as seen in the news sequence.

TABLE 1. PSNR FOR FIRST 100 FRAMES PER SEQUENCE.

	Frame Copy	Proposed
City	**31.6783**	29.5484
Coastguard	34.1663	**34.1839**
Container	34.8931	**35.0891**
Crew	21.8288	**23.2015**
Football	**34.8137**	34.7956
Foreman	16.6554	**18.3011**
Mobile	13.0802	13.0802
News	15.3796	**19.8205**
Soccer	34.4975	34.4975

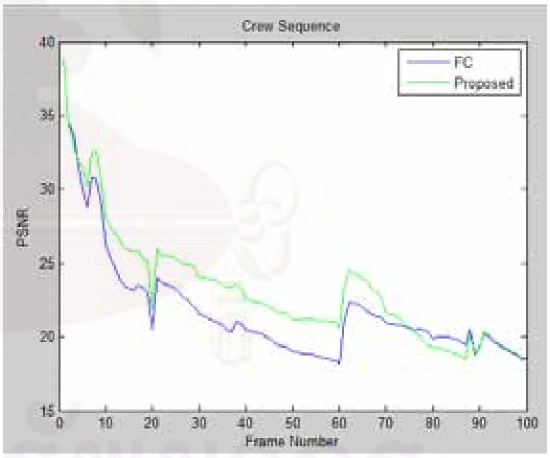


Figure 2. PSNR plot for Crew sequence.

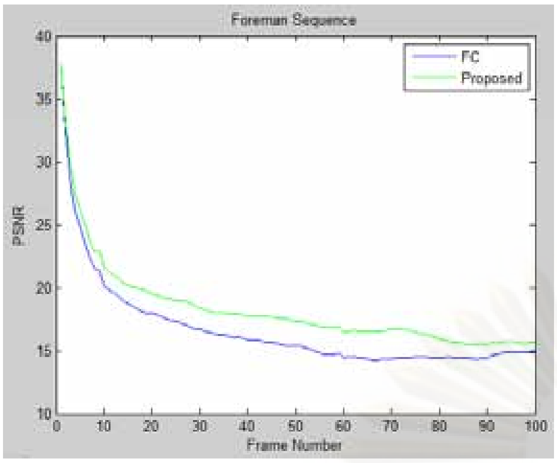


Figure 3. PSNR Plot for Foreman sequence

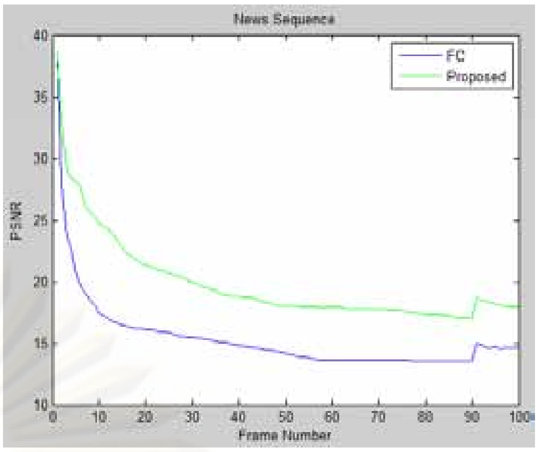


Figure 4. PSNR Plor for News Sequence

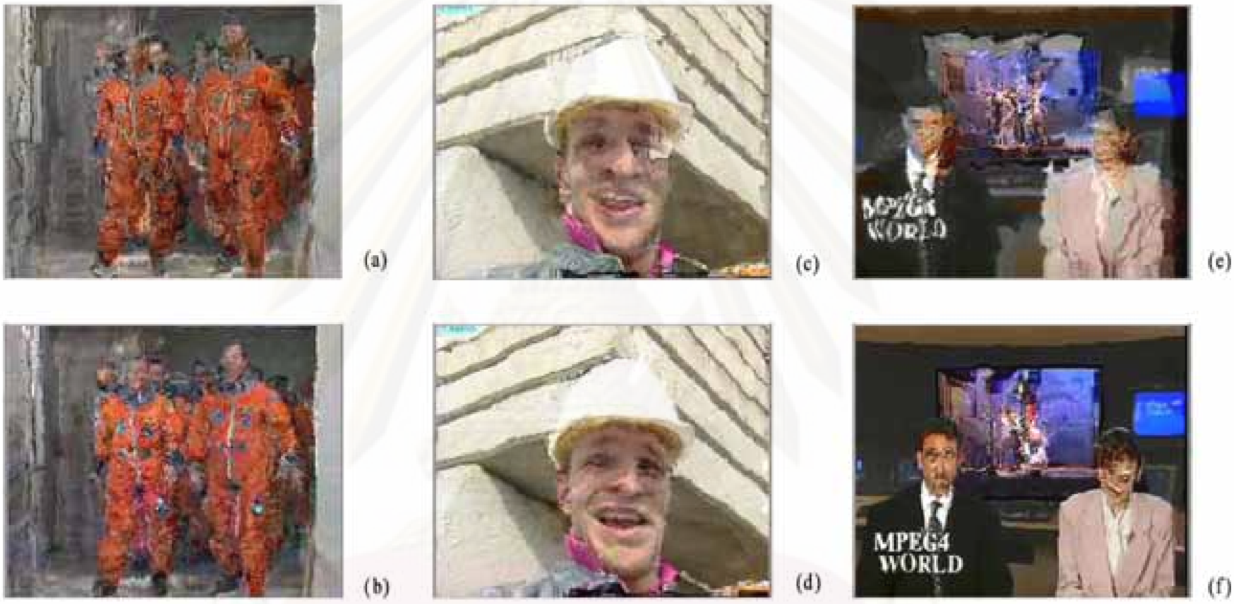


Figure 5. Subjective evaluation of the proposed algorithm and Frame Copy method. The frames above (a, c and e) were concealed using the frame copy method. The frames below ((b, d and f) were concealed using the proposed algorithm.

Fig. 5 shows sample frames from the Crew, Foreman, and News sequences. In the Crew sequence, we can see that the hands and feet of the people are better concealed in our method. Also, in FC, the faces of the people are completely distorted. In the Foreman sequence, we can observe that the blocking artifacts from our method are less compared to FC. The News sequence also suffers from severe distortion after concealing with Frame Copy. Our method is able to adequately estimate the lost motion information, hence preserving the integrity of the reconstructed frame.

In other sequences, the improvement of our method over the FC method is marginal. Upon observation, those sequences contained several macroblocks encoded in *intra* mode, hence no associated motion vector is needed for that macroblock.

CONCLUSION AND FUTURE WORK

We have shown that by exploiting the temporal redundancy in successive video frames, it is possible to generate a satisfactory estimate of the lost motion information and reconstruct the lost frame. More importantly, the proposed method is able to recover from base layer losses. Further improvement may be achieved by using a smaller block size in motion estimation, so as to further reduce blocking artifacts. Spatial error concealment may also be applied after the motion compensation step in order to improve the quality of the output frame. However, as also reflected in the results, our method is effective only in high motion sequences. The motion estimation method may also be applied adaptively, as it may not be needed in case of intra-coded macroblocks.

ACKNOWLEDGEMENT

This research has been supported in part by the Collaborative Research Project entitled Wireless Video Transmission, JICA Project for AUN/SEED-NET, Japan.

REFERENCES

H. Schwarz and M. Wien, "The Scalable Video Coding Extension of the H.264/AVC Standard: Standards in a nutshell. *IEEE Signal Processing Magazine*, vol. 25, pp. 135-141, Mar. 2008.

Han S. Error Propagation in Scalable Video Coding, Avaliable from http://ise0.stanford.edu/class/ee392j/projects/han_report.pdf

Y. Chen, K. Xie, F. Zhang, P. Pandit, and J. Boyce, "Frame Loss Error Concealment for SVC", *Journal of Zhejiang University SCIENCE A*, Vol. 7, No. 5, pp. 677-683, 2006.

J.H. Kim, M.C. Hwang, H.Y. Yang; S.J. Ko; Morales, A.W., *"Frame Error Concealment Technique Using Adaptive Inter-Mode Estimation for H.264/AVC,"* Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on , vol., no., pp.1-2, 9-13 Jan. 2008

M. M. Ghandi, M. Ghanbari, "Error concealment for SNR scalable video coding," *Signal Processing: Image Communication*, Vol. 21, Issue 2, pp. 91-99, February 2006.

Schwarz, H.; Marpe, D.; Wiegand, T., "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.9, pp.1103-1120, Sept. 2007

VITAE

Simon Jude Que Lam is a native of the Philippines. He was born on the 27^{th} of January, 1987. His bachelor's degree was earned in De La Salle University, Manila in September 2007, where he majored in Electronics and Communications Engineering. For his undergraduate thesis, he was involved in research focusing mainly on enhancement of medical images. He briefly taught laboratory subjects in De La Salle University and is also a licensed Philippine Electronics Engineer. He entered Chulalongkorn University for his Master’s degree in Electrical Engineering through the support of AUN/SEED-Net and JICA. He belongs to the Video Technology Research Group, focusing his studies on H.264/AVC and the application of error concealment in the aforementioned standard.