Scalable Video Coding with Invertible Sampling Lattice Conversion

A Dissertation
in
Information Science and Engineering
Submitted to Graduate School of Science and Technology
Niigata University
in
Partial Fulfillment of the Requirements
for the Degree of
Doctor of Engineering

Takuma Ishida

March 2006

Abstract

The goal of this study is to develop non-separable multi-rate systems integrated with invertible video format conversion in order to offer a spatio-temporal resolution control function as a new option for scalable video coding.

Chapter 1 describes a background of this study. As a comprehensive consequence of the explosive growth of Internet, great advances in hardware technologies and software developments, a lot of new multimedia applications are emerging rapidly. Although the storage capacity of digital devices and the bandwidth of networks are increasing constantly and rapidly, video compression still plays an essential role in these applications. In the applications of video delivery services such as video streaming over Internet and mobile communication, it is important to offer dynamic adaptability to a diversity of different networks, terminals and video formats with a seamless bitstream that conveys the video content. In this situation, a scalable function provides adaptive control capability of image quality (i.e. SNR), temporal rate, and picture size (resolution). Scalable video coding is an attractive way to universal and seamless access to video contents with dynamic controllability that meets different requirements in different networks, terminals, formats and displays including a TV and a PC.

Given these backgrounds, one of the hard problems in scalable video coding is a format conversion between interlaced and progressive scanning. The motivation of this study is to deal with a coding system using a new class of deinterlacing, which is referred to as invertible deinterlacing, for motion picture coding. Deinterlacing is a kind of sampling rate conversion of video signals. Usually it is defined as the sampling rate up-conversion of an interlaced scan signal to its progressive scan counterpart by interpolating the missing lines in every field. The deinterlaced signal will have twice the vertical sampling rate to the original. The invertible deinterlacing technique is defined in a new form different from the conventional deinterlacing techniques. In applications of scalable video coding, twice the video information data to the original is undesirable before encoding. Given an interlaced scanning video signal, the invertible conversion produces its corresponding progressive scanning video signal, where both of these two signals have the same sampling density, while maintaining the perfect reconstruction property. The objective of this study is to develop coding systems equipped with new spatio-temporal scalability. In the following chapters, several multirate systems with the invertible conversion are discussed so as to find a better solution.

Chapter 2 presents a lossy implementation scheme of Motion-JPEG2000 (MJP2) integrated with invertible deinterlacing. The invertible deinterlacing technique is developed in order to suppress comb-tooth artifacts which are caused by field interleaving for interlaced scanning videos and affect the reconstructed picture quality in scalable frame-based codecs such as MJP2. The technique is characterized by two features. The sampling density is preserved and image quality is perfectly recovered by an inverse process. When no codec is placed between the deinterlacer and inverse process,

the original video is perfectly reconstructed. Otherwise, it is almost completely recovered. An application scenario of this invertible deinterlacer is also presented for enhancing the sophisticated SNR scalability in the frame-based MJP2 coding. The proposed system suppresses the comb-tooth artifacts at low bitrates, while enabling the quality recovery through its inverse process at high bitrates within a standard bitstream format. The main purpose of this chapter is to present a system that yields high quality recovery for an MJP2 codec in lossy compression. We demonstrate that our invertible deinterlacer can be embedded into the discrete wavelet transform (DWT) employed in MJP2. As a result, the energy gain factor to control rate-distortion (R-D) characteristics is compensated for an optimal compression. Finally, the performance evaluation of image recovery by the proposed method is discussed. Several test sequences and filter characteristics are investigated through the simulation results.

Chapter 3 describes a lossless compression technique of MJP2 integrated with the invertible deinterlacers. The proposed system suppresses comb-tooth artifacts in frame pictures generated from interlaced pictures, and is applicable to both of frame and field based-displaying applications with a single code-stream. It is easily imagined that those frame pictures are frequently edited and recompressed regardless of being used in professional or consumer applications. Usually, it is required to preserve all accurate information during such repetitive processes. The reversible, i.e. lossless, conversion to reconstruct interlaced images from deinterlaced code-streams is therefore of significant interest. Lossless compression is desirable in these applications where images are likely to be edited and recompressed reversible times, whereas the accumulation of errors by multiple lossy compressions can grow to an intolerable level. This chapter describes an exact lossless implementation technique. The proposed implementation is realized by integrating the deinterlacer into MJP2 codecs and modifying the header information for appropriate standard decoding. Some simulation results show that the comb-tooth suppression capability is kept at low bitrates with standard MJP2 decoders.

Chapter 4 discusses motion compensated three-dimensional (3-D) filter banks providing a new tool of a spatio-temporal resolution control for scalable video coding. There are recent developments in scalable video coding, and most of them are yet based on 3-D wavelet transform with motion compensation. To achieve the function of frame-rate and spatial resolution scalabilities, motion compensated temporal filtering (MCTF) through a lifting wavelet transform currently attracts many researchers as an effective temporal decomposition tool. To provide the video format control between the progressive and interlaced scanning, two different structures with a spatio-temporal split process are presented as a unique spatio-temporal scalability. Unlike other filter banks, the proposed systems are constructed in a unique way to multi-dimensional systems. Two kinds of interlaced video formats are suggested: the vertical-temporal (VT) quincunx and face-centered-orthorhombic (FCO) format.

At first among them, perfect reconstruction (PR) deinterlacer banks are suggested. The proposed system separates a progressive video into a pair of different progressive videos of a half frame rate. It is novel from the viewpoint of filter banks where interlaced videos are given as intermediate data during analysis and synthesis process. Unlike the conventional filter banks, our systems are constructed in a way unique to multidimensional systems by using invertible deinterlacers which we have proposed. This technique offers a functionality by which a reconstructed base layer itself provides both of half-rate progressive and interlaced videos, and an additional enhancement layer, which has a prediction error in high-pass subbands, improves the spatio-temporal resolution.

Secondly, a novel non-separable motion compensated spatio-temporal filter (MC-

STF) is proposed. This technique can provide the video format control between progressive and interlaced scanning for spatial-temporal scalable video coding as well as deinterlacer banks. This technique corresponds motion compensated temporal filtering, which has an important role as a component of 3D-DWT for exiting scalable video coding techniques. As a new tool, we proposed a motion compensation technique for non-separable sampling lattices in terms of spatial-temporal decomposition. The proposed method has a different structure with motion compensated field prediction from the deinterlacer banks. In some experimental results, we show the significance of these proposed techniques.

Finally in Chapter 5, conclusion of each chapter is drawn after an entire summary of this study.

Acknowledgements

I would like to take this opportunity to express my deep gratitude to those who supported me in conducting this study and in writing this thesis. First and foremost, I would like to thank advisory professors, Dr. Shogo Muramatsu and Dr. Hisakazu Kikuchi for their guidance, patience and support throughout my undergraduate and graduate research activities with every aspect. But for their encouragement, I might have thrown up completing this research. I would like to thank the committee members of this thesis, Prof. Masanobu Yamamoto and Prof. Osami Sasaki, Niigata University. Their valuable comments were very helpful for the progress in this thesis. I would like to express my thanks to Mr. Tetsuro Kuge, NHK Science and Technical Research Laboratories for his helpful comments and for his constructive suggestions. I would also like to thank Dr. Shigenobu Sasaki, Niigata University, and Dr. Jie Zhou, presently, Nanjing Institute of Technology, for a lot of valuable advises and their effort. Their contribution to this thesis was invaluable. I would like to express my deep thanks to Japan Society for the Promotion of Science for granting the honor of JSPS Research Fellowship with a valuable support. I would like to thank the Ministry of Education, Culture, Sports, Science and Technology for its support by the Grand-in-Aid for Scientific Research (JSPS Fellows) No.16 · 5404. I would additionally like to thank Dr. Yoshito Abe, Industrial Research of Niigata Prefecture, Dr. Naoki Mizutani, Kodak Digital Product Center Japan Ltd., Dr. Satoshi Hasebe,Wireless and Visual Communications Co., Ltd., Mr. Minoru Hiki, Niigata University, and Mr. Jun Uchita, Fujitsu LSI Solution Ltd., for their encouragements. Without their helpful comments and encouragement, this thesis would not meet its completion. Although there have appeared no names here, I thank to all individuals in Department of Electrical and Electronic Engineering and Graduate School of Science and Technology, Niigata University, my friends and relatives for their generous support. Finally, I would like to my special appreciation to my family. They have been giving me constant support and encouragement. I enjoyed my study throughout the period.

Takuma Ishida

Niigata University
March 2006

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Purpose of This Study

The objective of this study is to deal with video coding systems providing the video format control between the progressive and interlaced scanning manners for scalable video coding.

As is widely observed, the convergence of television and computers including mobile computing terminals is in progress. In the broadcasting community the interlaced scan format is used in TV, whereas a PC uses progressive scan formats or picture frames [1–3]. The telecommunication community seems to be a big user of video coding standards and multimedia contents involved these video formats.

In resent years, with the explosive growth of Internet, the great advances in hardware technologies and software developments, a lot of new multimedia applications are emerging rapidly. Although the storage capability of the digital devices and the bandwidth of the networks are increasing constantly and rapidly, video compression still plays an essential role in these applications. With facilitating widely the equipment of network infrastructures such as broadband network (such as DSLs), mobile communication services and wireless local area network (LAN), it is becoming important to offer a flexible service and more functionality that achieves an adaptive control providing the bit-rate and the quality in according to the bandwidth dependent on different networks and the capability of various terminal devices. In the applications of video delivery such as video streaming over heterogeneous network environments, the video coding techniques have to offer dynamic adaptation to a diversity of different networks, terminals and video formats with a seamless bit-stream. Therefore, in this situation, the video coding techniques require not only high coding efficiency, but also fine grain functionalities and flexibility to the end-users. Scalable functions that offers the image quality (SNR), temporal rate, picture size (resolution) can be controlled adaptively. Obviously, fine grain scalability is an efficient way to get universal and seamless access to the video contents with adaptation to various terminal, access conditions and display such as TVs, PCs, and mobile devices.

Once given this background, one of the challenging problems is a format conversion between interlaced and progressive scanning. Not only the problem is deeply involved with the existing television systems and video camera and displaying systems, but it is of significant relevance to video coding schemes, since scalable video coding can accelerate the convergence as well as transcoding in advanced informa-

tion/telecommunication environment.

As the border of these communities fades away more and more, it is becoming important to offer the format conversion between progressive and interlaced scanning manners. Deinterlacing is necessary to convert the scanning format of motion pictures from interlaced to progressive scan [1–3]. Historically, deinterlacing has been used to avoid the aliasing and flicker artifacts for better visual quality in broadcast TV signals, and to convert formats between different schemes, for instance, between NTSC and PAL as well as from NTSC to VGA. Deinterlacing is a kind of sampling rate conversion of video signals [4, 5]. Usually it is defined as the sampling rate up-conversion of an interlaced scan signal to its progressive scan counterpart by interpolating the missing lines in every field. The deinterlaced signal will have twice the vertical sampling rate to the original. The conventional deinterlacing is used as post-filtering in decoder sides to make moving objects clearer and smoother in terms of visual view point for progressive display systems.

In contrast, the problem to be addressed in this thesis is different from the conventional deinterlacing. If the applications of scalable coding systems as pre-filtering are assumed, twice the video information data to the original is undesirable before encoding. Given an interlaced scanning video signal, we are trying to find its progressive scanning video signal, where both of these two signals have the same sampling density [12, 13].

The key of this study is to deal with a new class of deinterlacing for motion picture coding that preserves the sampling density in interlaced original videos and deinterlaced videos and thus provides the reversible conversion between interlaced and progressive scan formats. In particular, low bit-rate decoding causes considerable degradation around the edges because of the loss of high frequency components. Let us assume an embedded scalable codec system such as MJP2. Although the pre-filtering before encoding removes the comb-tooth artifacts, it degrades the spatial resolution in high bit-rate decoding. On the other hand, the post-filtering after decoding could solve the problem in the low bit-rate decoding without any loss of resolution for high bit-rate decoding at cost of extra operations that is unfavorable for low bit-rate decoders.

The contribution of the study is summarized as follows:

1. To have given a coding system with the invertible deinterlacer for intra-based coding application

2. To have provided video format control between progressive and interlaced scan formats using non-separable multirate systems

The first point is subjected to the implementation issue, and a simple but effective implementation developed in connection to an encoding/decoding system for both of lossy and lossless compression. In lossy compression, for the suggested application scenario of a MJP2 codec with the deinterlacer, the proposed gain compensation can achieve the improvement of the coding efficiency, while keeping the comb-tooth suppression. In lossless compression, this thesis describes a way to resolve a problem caused by discarding fractions produced with the deinterlacer, and its lossless implementation can be realized MJP2 codec integrated with the invertible deinterlacer.

The second point is to have developed non-separable multirate systems with motion compensation technique to provide a new option i.e. video format control between rectangular and non-rectangular sub-sampling. Two types of the structures with spatio-temporal split process are proposed to achieve spatio-temporal decomposition as new

scalable functionalities. One is the implementation with the invertible deinterlacer. The other is the configuration applying field prediction with motion compensation. The proposed systems separate a progressive video into two different progressive videos of a half frame rate and enable us to handle interlaced videos as intermediate data during analysis and synthesis process. In the thesis, it is clarified that these proposed techniques can bring a new spatio-temporal scalability that provides the video format control between the progressive and interlaced scanning manner. In addition, the proposed techniques can suppress PSNR fluctuations generated using the lifting-based MCTF technique without the update step.

1.2 Overview

The rest of this thesis is organized as follows:

Chapter 2 presents an implementation scheme of Motion-JPEG2000 (MJP2) integrated with invertible deinterlacing. The invertible deinterlacing technique is a new class of format conversion between interlaced and progressive scanning formats. This deinterlacer enables us to suppress comb-tooth artifacts which are caused by field interleaving for interlaced scanning videos and affect the quality of scalable frame-based codecs such as MJP2. The technique has two features, where sampling density is preserved and image quality is recovered by an inverse process. When no codec is placed between the deinterlacer and inverse process, the original video is perfectly reconstructed. Otherwise, it is almost completely recovered. An application scenario of this invertible deinterlacer is suggested for enhancing the sophisticated SNR scalability in the frame-based MJP2 coding. The proposed system suppresses the comb-tooth artifacts at low bitrates, while enabling the quality recovery through its inverse process at high bitrates within the standard bitstream format. The main purpose of this chapter is to present a system that yields high quality recovery for an MJP2 codec in lossy compression. It is demonstrated that the invertible deinterlacer can be embedded into the discrete wavelet transform (DWT) employed in MJP2. As a result, the energy gain factor to control rate-distortion (R-D) characteristics can be compensated for optimal compression. Finally, the performance evaluation of image recovery with the proposed method is discussed. And, several test sequences and filter characteristics are investigated through the simulation results.

Chapter 3 describes lossless compression of MJP2 integrated with the invertible deinterlacers. For the proposed application scenario in the preceding chapter, our system suppresses comb-tooth region in frame pictures generated from interlaced pictures and is applicable to both of frame and field based-display applications with single code-stream. It is easily imagined that those frame pictures are frequently edited and recompressed regardless of being used in professional or consumer applications. Usually, it is required to preserve all accurate information during such repetitive processes. The reversible, i.e. lossless, conversion supporting to reconstruct the interlaced images from the deinterlaced code-stream is therefore of interest. Lossless compression is desirable in applications where the images are to be extensively edited and recompressed, whereas the accumulation of errors by multiple lossy compression might become unacceptable. In this chapter, it is proposed an exact lossless implementation technique by integrating the deinterlacer into MJP2 codecs and modifying the header information for appropriate standard decoding. Some simulation results are demonstrated to show that the comb-tooth suppression capability can be kept at low bitrates with standard MJP2 decoders.

Chapter 4 discusses motion compensated three-dimensional (3-D) filter banks as a new tool of a spatio-temporal resolution control for scalable video coding. For recent developments in scalable video coding, most of them are based on 3-D wavelet transform with motion compensation. To achieve the function of frame-rate and spatial resolution scalabilities, motion compensated temporal filtering (MCTF) through lifting wavelet transform currently attracts many researchers as an effective temporal decomposition tool. To provide the video format control between the progressive and interlaced scanning manner as an unique spatio-temporal scalability, two different structures with a spatio-temporal split process are presented. Unlike other filter banks, the proposed systems are constructed in a way unique to multi-dimensional systems. Two kinds of interlaced video formats are suggested: the vertical-temporal (VT) quincunx and face-centered-orthorhombic (FCO) format.

Firstly, perfect reconstruction (PR) deinterlacer banks are suggested. The proposed systems separate a progressive video into two different progressive videos of a half frame rate and are novel from the viewpoint of filter-banks in that interlaced videos are given as intermediate data during analysis and synthesis process. Unlike the conventional filter banks, our systems are constructed in a way unique to multidimensional systems by using invertible deinterlacers which we have proposed before. This technique offers a functionality by which reconstructed a base layer by itself provides both of half-rate progressive and interlaced videos, and adding an enhancement layer, which have the prediction error in high-pass subband, improves the spatio-temporal resolution.

Secondly, a novel non-separable motion compensated spatiotemporal filter (MC-STF) is proposed. This technique can provide the video format control between progressive and interlaced scanning for spatial-temporal scalable video coding as well as deinterlacer banks. This technique corresponds motion compensated temporal filtering, which has an important role as the component of 3D-DWT for the exiting scalable video coding techniques. As a new tool, it is proposed motion compensation technique for non-separable sampling lattices in terms of spatial-temporal decomposition. This proposed method has different structure with motion compensated field prediction from the deinterlacer banks. Some experimental results show the significance of these proposed techniques.

Finally, in Chapter 5, each conclusion of each chapter follows after a whole summary of this study.

Chapter 2

Lossy Compression for Interlaced Video Sequences

The main purpose of this chapter is to present a system that yields high quality recovery and suppress annoying artifacts for scalable intra-codecs in lossy compression for interlaced video sequences. An implementation scheme of Motion-JPEG2000 (MJP2) integrated with invertible deinterlacing is discussed. The invertible deinterlacing technique is developed in order to suppress comb-tooth artifacts which are caused by field interleaving for interlaced scanning videos and affect the quality of scalable frame-based codecs such as MJP2. The technique has two features, where sampling density is preserved and image quality is recovered by an inverse process referred to as reinterlacer. When no codec is placed between the deinterlacer and inverse process, the original video is perfectly reconstructed. Otherwise, it is almost completely recovered. An application scenario with this invertible deinterlacer is presented for enhancing the sophisticated SNR scalability in the intra-frame-based MJP2 coding. The invertible deinterlacer can be embedded into the discrete wavelet transform (DWT) employed in MJP2. As a result, the energy gain factor to control rate-distortion (R-D) characteristics can be compensated for optimal compression. Simulation results show that the recovery of image quality is improved after reinterlacing, while keeping the suppression of comb-tooth artifacts.

2.1 Introduction

Interlaced scanning and progressive scanning are known as recording and displaying formats for motion pictures [1, 2, 6, 7]. In the television broadcasting community, the video standards, e.g. some of BT.601 and SMPTE 295/296M (NTSC, PAL and SECAM) signals, are based on the interlaced scanning format, whereas currently, in the visual communication community including mobile communication, videophone, and video on Internet, video contents and video coding standards are developed in the progressive scanning format.

Video coding techniques are classified into two categories according to the way of treatment in temporal correlation: intra-coding and inter-coding. The inter-coding is in general superior to the intra-coding in terms of coding efficiency, and is well observed in widely accepted standards MPEG1/2/4 and H.26x [8]. These standards adopt a hybrid of the transform-based and motion-compensated coding techniques. On the other

hand, the intra-coding technique also receives strong supports in the field of content editing, visual information retrieval and communication with robust error resilience. No motion compensation is employed, and every frame is encoded independently so that handling of every frame is quite easy and errors in a frame never propagate to the others.

One of the representative intra-coding standards is Motion-JPEG2000 (MJP2) as Part 3 of the JPEG2000 (JP2) image coding standard, which is based on the baseline JPEG2000 Part 1 [24–28]. Recent explosive growth requires various kinds of scalable functionality in audio and visual coding techniques as well as high compression performance at preferable quality. To cope with these requirements, JP2/MJP2 excellently covers the SNR, spatial, frequency, and component dominant (color) scalabilities. The SNR scalability enables us to obtain a video content at an arbitrarily specified bitrate with moderate quality by partial decoding of a bit stream that has been encoded at a higher bitrate. This functionality is quite useful when a viewing user wants to control the balance between quantity and quality. This chapter is concerned with this SNR scalability of MJP2 for interlaced scanning video contents.

With regard to interlaced sources, there are two classes in the way of managing field pictures: field-based coding and frame-based coding. The former encodes two successive fields of a frame separately and the latter encodes every frame as a single picture. Actually, MJP2 defines both of field- and frame-based coding modes. The article [29] describes that the relative rate-distortion (R-D) performance of these two coding modes highly depends on the properties of given sources. For a picture that shows strong spatial consistency across adjacent rows, the frame-based mode is preferable to the field-based mode. For example, it is the scene containing abundant detail where a camera has little action and objects in the sequence move slowly and occupy small areas. In contrast, the field-based mode is favorable for a video sequence including a fast camera action as well as fast moving objects covering large areas in a scene with little detail. From this observation, it is recommended to introduce picture-adaptive frame/field (PAFF) coding, which adaptively switches the coding mode depending on the characteristics of a video sequence. The followings are also the cases where the frame-based mode is advantageous.

- A sequence contains small moving objects with a still dominant background or slowly moving dominant foreground objects.
- A simple encoding structure is preferred instead of introducing a mode selector.

The frame-based mode requires a field interleaving (or field-merging) process so as to generate a frame picture from two successive fields. Unfortunately, this process causes horizontal comb-tooth artifacts at edges of moving objects [10]. In other words, frames generated through field interleaving suffer from a tricky problem because of vertical-temporal aliasing or a vertical-temporal offset between two successive fields in the same frame [2,6]. In the case of transform-based coding such as MJP2, if the vertically high-frequency components involved with comb-tooth artifacts are quantized, the quantization errors appear annoyingly and become perceivable to the naked eye, especially at low bitrates. This fact implies that some remedy for the comb-tooth artifacts is required for making the most of the SNR scalability in MJP2 for interlaced scanning video.

So far, two different techniques have been studied for solving the comb-tooth artifacts in motion picture coding. One of them is the motion-compensated (MC) deinterlacing approach [2,3,30,31], and the other is the linear pre-filtering approach [2,3,10].

The MC approach shown in the article [31] takes account of an optimal quantization gain with respect to the sensitivity of visual perception. In particular, the MC technique seems more attractive for inter-coding systems such as MPEG2 because computational resources for motion compensation can be shared. Since missing lines in every field are interpolated along the motion trajectory, the MC approach performs well for both of still and moving parts. On the other hand, this approach requires a huge amount of processing and the transmission of motion vectors to reconstruct interlaced pictures. Thus, it is not necessarily appropriate to apply the MC approach to intra-coding such as MJP2. In addition, when the motion estimation fails, the error grows perceptually large.

Recently, a simple MJP2 system with a pre-filter was proposed by Kuge [10] to suppress these artifacts. This technique was demonstrated to be effective for decoding at a target bitrate. It was especially effective for low bitrate decoding[1]. However, since the scalable functionality hides decoding bitrates from encoders, nobody knows how much the decoding bitrate is while encoding. When the decoding bitrate is higher than a fixed bitrate targeted by involved filtering, the quality of a picture is degraded by the effect of lowpass filtering. Such behavior is not desirable with respect to the SNR scalability.

To solve this problem, in previous works, we developed an invertible deinterlacing technique[2] with sampling density preservation as a preprocess for scalable frame-based coding [12, 13]. Then, we designed several deinterlacing filters and suggested a scenario of using a deinterlacer prior to a frame-based coding, where the deinterlacer was used to enhance the scalability of MJP2 [14–16]. With this technique, we could suppress comb-tooth artifacts at low bitrates while guaranteeing the quality recovery through the inverse process at high bitrates. Such recovered pictures at high bitrates, however, have less peak signal-to-noise ratio (PSNR) than that obtained by the conventional process with field interleaving. For example, the difference is around 2 dB when decoding 8-bit grayscale sequence *Football* in the MJP2 bitstreams at 2.0 bpp.

The purpose of this chapter is to present a system that improves the quality in decoding at high bitrates from the direct application of invertible deinterlacing to MJP2, while suppressing the comb-tooth artifacts at low bitrates. To achieve this purpose, we firstly analyze the influence of an invertible deinterlacer to the error energy distribution in the wavelet transform domain of JP2, and propose an energy gain compensation technique for giving optimal R-D control. The energy gain factor is computed from the L2 norm of the inverse filter coefficients for the relevant subband [25, 28]. Thus, the problem can be reduced to know how the inverse filter coefficients change due to the invertible deinterlacer. As a previous work, we have already shown that invertible deinterlacing can be embedded into the discrete wavelet transform (DWT) of JP2 in some special cases [19], where the computational costs were only concerned. In this chapter, it will be shown that this fact can be utilized for analyzing the influence of the deinterlacing and that a mismatching energy gain factor can be compensated, while suppressing the effect of comb-tooth artifacts within the standard bitstream format. Several simulation results show that this optimal recovery method is superior to the

[1] In this chapter, based on Table1.1 in the text [28], we refer to bitrates higher than 1 bpp as high bitrates, and those lower than 0.25 bpp as low bitrates.

[2] Term 'invertible' is used to indicate the capability of a perfect reconstruction where an inverse system analytically exists, and to distinguish the technique from the existing deinterlacing popularly used in advanced TV receivers [2]. In fact, the design problem of deinterlacing filters can be reduced to a design problem of multi-dimensional two-channel maximally-decimated filter banks [13]. The lossless or reversible property is not necessarily assumed.

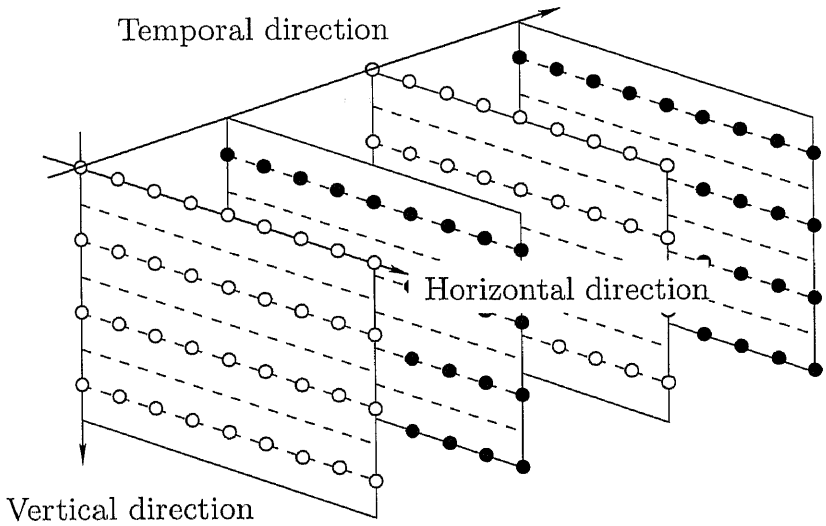


Figure 2.1: Interlaced scanning, where open and closed circles are sample points in top and bottom fields, respectively.

default gain approach through some video sequences of different characteristics.

2.2 Review of Invertible Deinterlacing

In previous work, we proposed a deinterlacing technique that preserves sampling density and has invertibility [12, 13]. Here, let us review this technique as a preliminary to the more detailed discussion that follows. Assume that input array $X(\mathbf{z})$ is given as shown in Fig.2.1.

2.2.1 Deinterlacing that Preserves Sampling Density

Figure 2.2 outlines the basic structure of deinterlacing that preserves sampling density, where $\mathbf{z}$ is a 3-D vector that consists of 3-D z-domain variables. The upsampler converts interlaced video signal array $X(\mathbf{z})$ into another array that is non-interlaced as shown in Fig. 2.2. Filter $H(\mathbf{z})$ that follows is a 3-D spatio-temporal filter, which suppresses comb-tooth artifacts. To preserve input array density, the processed video signal array is finally downsampled in the temporal direction (Fig. 2.2) to obtain deinterlaced array $Y(\mathbf{z})$. This process can be regarded as a generalization of field interleaving.

2.2.2 Reinterlacing (Inverse Process)

The deinterlaced array $Y(\mathbf{z})$ is encoded, transmitted and then decoded frame by frame in a frame-based codec system. The interlaced video source is expected to be reconstructed from frame-based pictures in some applications. To achieve this, we introduced a process of inverse conversion, i.e. *reinterlacing*. Figure 2.3 illustrates the

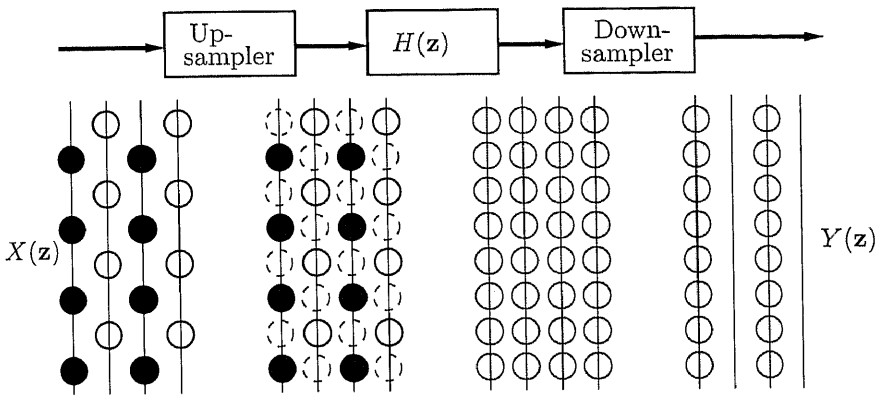


Figure 2.2: Basic structure of deinterlacing with preservation of sampling density, where only vertical-temporal plane is shown. The dotted and gray circles indicate sample points of inserted zero values and deinterlaced lines, respectively.

basic structure. Array $Y(\mathbf{z})$ is firstly upsampled in the temporal direction, filtered with $F(\mathbf{z})$ and then downsampled (Fig. 2.3). We have shown how input array $X(\mathbf{z})$ is perfectly reconstructed from deinterlaced array $Y(\mathbf{z})$ when $H(\mathbf{z})$ and $F(\mathbf{z})$ satisfy a set of perfect reconstruction conditions in the articles [12] and [13]. If the conversion system does not suffer from any rounding error, it is theoretically possible to recover the original picture in perfect.

2.2.3 Application Scenario

An application scenario can be developed on the suggested codec system [14, 15] outlined in Fig. 2.4. It offers the compatibility and coexistence of interlaced and progressive scanning video formats and also provides an extension of decoding options at front-end receivers. An encoded single bit-stream is delivered to a decoder, and front-end users select the most desirable and optimum decoding among various decoding options depending on their displaying terminals and available transmission bitrates. It can be referred to as single-source encoding/multiple-display decoding. A difference from our previous work is that the encoder in this work is optimally designed for the reinterlacer. Its design affects just a few parametric values in MJP2 and gives no critical changes in the encoding procedure. It should be noted that nobody is assumed to know exact decoding conditions during the process of encoding.

Figure 2.4 shows the block diagram of the process flow for the frame-based mode. For the field-based mode, we assume no special process. As illustrated in Fig. 2.4, decoding can be roughly divided into four cases. Note that there is no clear border of low and high bitrate decoding. Thus, an important point here is that both of interlaced and frame-based video contents are obtained from the same compressed bitstream on demand. The invertible deinterlacer ensures better low bitrate decoding because comb-tooth artifacts are suppressed. The decoded signals are displayed on a monitor directly or after passing through a reinterlacer depending on whether the displaying monitor is progressive or interlaced, respectively. The reinterlacing for low bitrate decoding

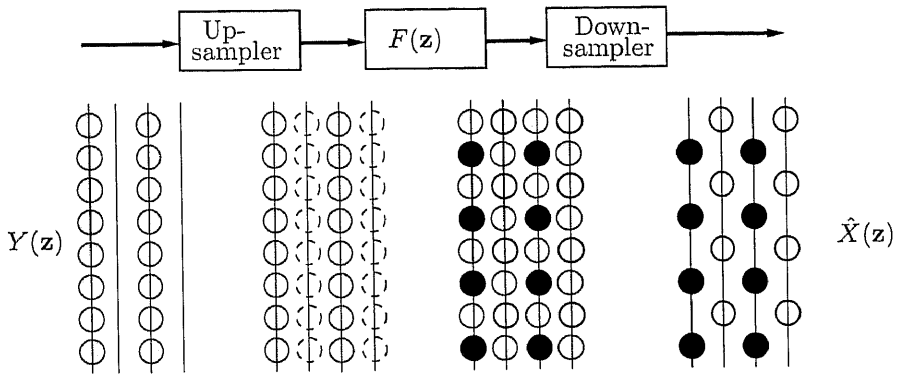


Figure 2.3: Basic structure of reinterlacing, where only vertical-temporal plane is shown. The open and closed, dotted, and gray circles indicate sample points of top line, bottom line, inserted zero value, and the other, respectively.

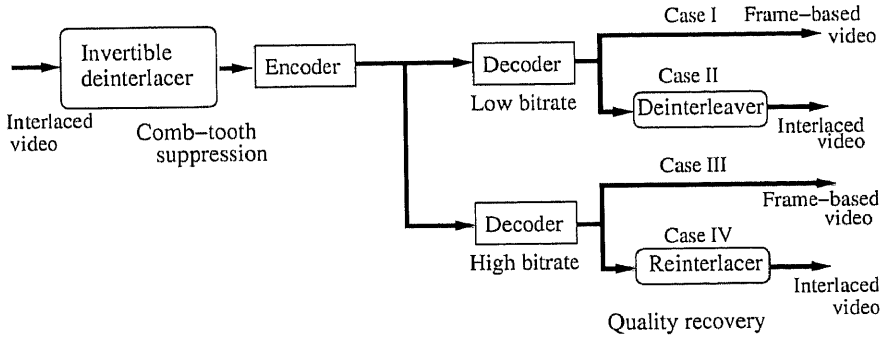
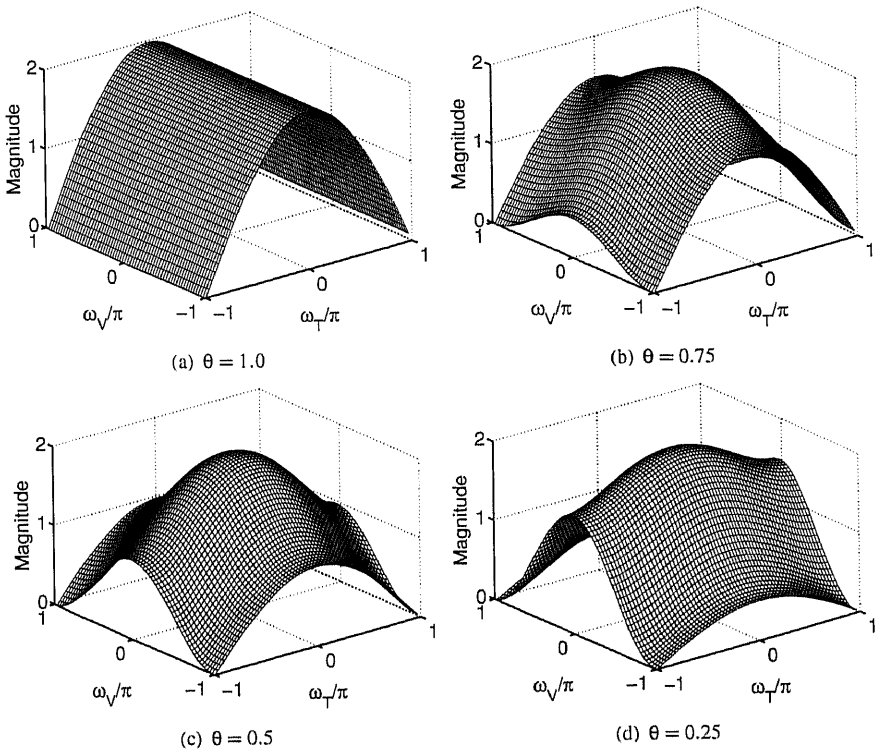


Figure 2.4: Frame-based coding/decoding system with an invertible deinterlacer.

may be replaced by deinterleaving that is the simplest reinterlacing to keep the comb-tooth suppression effect. For high bitrate decoding of interlaced video, the reinterlacer guarantees a quality recovery in terms of spatio-temporal resolution. In some cases, for example on advanced TV receivers, the reinterlaced video would further be processed to yield progressive video through another line-doubling deinterlacer [2], where the choice of the technique is beyond the scope of this thesis.

For the system, deinterlacing and reinterlacing do not affect any encoding/decoding functionalities except for a few gain parameters, and all scalabilities equipped with MJP2 are still effective.



(a) $\theta = 1.0$ (b) $\theta = 0.75$

(c) $\theta = 0.5$ (d) $\theta = 0.25$

Figure 2.5: Frequency magnitude responses of deinterlacing filter $H(\mathbf{z})$.

2.2.4 (3+1)-tap Deinterlacing Filter

A (3+1)-tap filter is an example of a deinterlacing filter presented in the articles [12] and [13]. The transfer functions of the filter and its dual are given as follows:

$$H(\mathbf{z}) = 1 + \theta z_T^{-1} + \frac{1-\theta}{2}(z_V + z_V^{-1}), \tag{2.1}$$

$$F(\mathbf{z}) = \frac{1}{\theta} z_T^{-1} \left\{ 1 + \theta z_T^{-1} - \frac{1-\theta}{2}(z_V + z_V^{-1}) \right\}, \tag{2.2}$$

where θ is a design parameter in range $0 < \theta \leq 1$. These filters have the following properties: i) normalized amplitude to maintain brightness, ii) regularity to avoid the checkerboard effect [20, 21] and iii) vertical symmetry to enable symmetric border extension [20, 22]. Additionally, all samples on their original position preserve their sample values.

The frequency characteristics of the filter can be changed between temporal, vertical-temporal (V-T), and vertical lowpass filters by selecting parameter θ as shown in Fig. 2.5. Particularly, when $\theta = 0.5$, all of the filter coefficients are given as powers of two. When $\theta = 1.0$, simple field interleaving is obtained. Note that this temporal filter ($\theta = 1.0$) has no capability of removing vertical high frequency components.

It has been verified that the filters can be computed through the in-place implementation as shown in Fig. 2.6 [19], where symmetric border extension is applied verti-

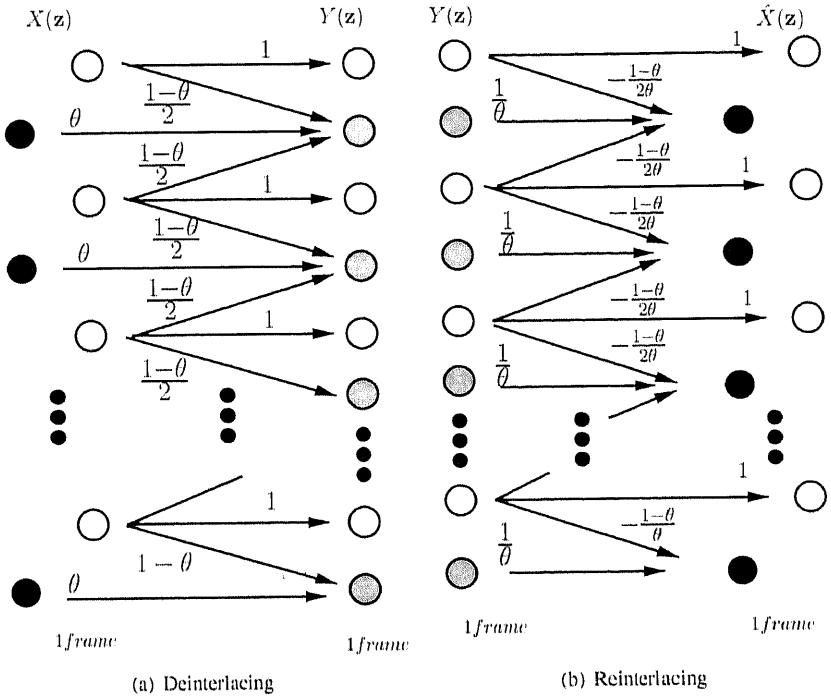


Figure 2.6: In-place computation of (3+1)-tap filters, where the symmetric extension is applied. The open, closed and gray circles denote samples of top, bottom and deinterlaced line, respectively, where only the vertical-temporal plane is illustrated.

cally. The open, closed and gray circles indicate pixels on the top line, the bottom line of $X(\mathbf{z})$, and the odd line of deinterlaced frame $Y(\mathbf{z})$, respectively. An odd line of deinterlaced frame pictures is obtained by the weighted sum of three lines with the values alongside the arrows. Note that this implementation is simple and efficient because redundant computations due to downsampling and upsampling are removed from the basic structures in Figs. 2.2 and 2.3.

Figure 2.7 shows the corresponding lifting structures. Each structure is composed of one prediction with simple Haar type two-tap filters. These lifting structures are equivalently applied to interleaved frame pictures in the vertical direction. It was verified that comb-tooth artifacts can be avoided for low bitrate decoding with this (3+1)-tap filter $H(\mathbf{z})$ of $\theta = 0.5$. Figures 2.8 (a) and (b) show pictures obtained by the field interleaving and invertible deinterlacing, respectively, where successive field pictures of a *Football* sequence (720 × 486 pixels, 8-bit grayscale) are used and the lossy mode of JP2 Part 1 (JJ2000 ver.4.1 [23, 26][3]) is applied to each frame. One can recognize that spot-like comb-tooth artifacts around the edges of moving objects have been suppressed by the invertible deinterlacing. Note that the field interleaving, on the other hand, has no capability of suppressing the comb-tooth artifacts.

The filtering effect of the deinterlacer sometimes gives us blurred pictures. Fortunately, we now know that it can be smartly removed by the reinterlacer if needed.

[3]We have fixed a bug of value δ in the forward 9/7 lifting transform module.

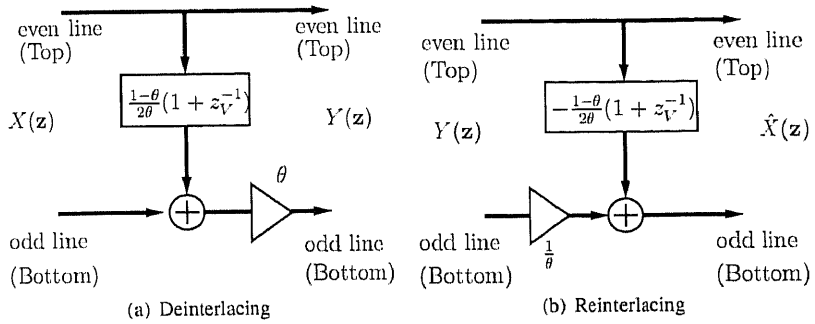


Figure 2.7: Lifting implementation of invertible scan format converters.

$$
\begin{aligned}
\mathbf{E}(z_V) &= \begin{pmatrix} E_{00}(z_V) & E_{01}(z_V) \\ E_{10}(z_V) & E_{11}(z_V) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{K} & 0 \\ 0 & K \end{pmatrix} \begin{pmatrix} 1 & \delta(1+z_V^{-1}) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ \gamma(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ \alpha(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{1-\theta}{2}(1+z_V^{-1}) & \theta \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{K} & 0 \\ 0 & K \end{pmatrix} \begin{pmatrix} 1 & \delta(1+z_V^{-1}) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ \gamma(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ (\alpha+\frac{1-\theta}{2})(1+z_V^{-1}) & \theta \end{pmatrix}
\end{aligned} \tag{2.3}
$$

$$
\begin{aligned}
\mathbf{R}(z_V) &= \begin{pmatrix} R_{00}(z_V) & R_{01}(z_V) \\ R_{10}(z_V) & R_{11}(z_V) \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ -\frac{1-\theta}{2\theta}(1+z_V^{-1}) & \frac{1}{\theta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\alpha(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & -\beta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ -\gamma(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} K & 0 \\ 0 & \frac{1}{K} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ -\frac{1}{\theta}(\alpha-\frac{1-\theta}{2\theta})(1+z_V^{-1}) & \frac{1}{\theta} \end{pmatrix} \begin{pmatrix} 1 & -\beta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} z_V^{-1} & 0 \\ -\gamma(1+z_V^{-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta(1+z_V^{-1}) \\ 0 & z_V^{-1} \end{pmatrix} \begin{pmatrix} K & 0 \\ 0 & \frac{1}{K} \end{pmatrix}
\end{aligned} \tag{2.4}
$$

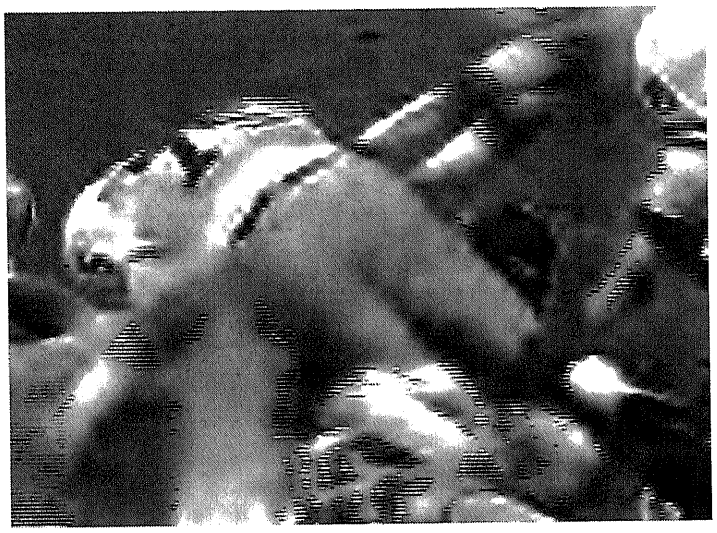
However, without any remedy, the recovered pictures show less PSNR than those obtained by field interleaving. The purpose of this work is to improve the quality of recovered images when MJP2 is used between the deinterlacer and the reinterlacer, while keeping the comb-tooth artifact suppression capability.

2.3 DWT Gain Compensation

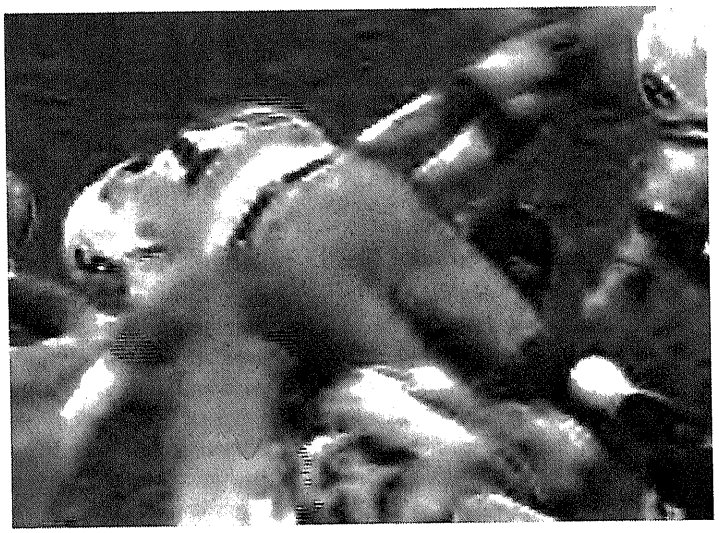
The quality of recovered pictures was not superior to that obtained by field interleaving. To improve this behavior, let us consider compensating the energy gain factor by embedding the invertible deinterlacer into DWT.

2.3.1 Energy Gain Factor

To minimize the distortion of a decoded picture in terms of the mean squared error (MSE) for a given bitrate, an energy gain factor is used in JP2. The energy gain factor is computed from the squared norm of the relevant subband's wavelet synthesis filter coefficients. According to the corresponding subband's energy gain factor, quantization errors of transform coefficients are weighted and the distortion of the decoded picture is estimated [25, 28]. The scenario assumes that deinterlaced images will be reinterlaced



(a) Conventional field interleaving



(b) Invertible deinterlacing $H(\mathbf{z})$ with $\theta = 0.5$

Figure 2.8: Magnified views of each picture decoded at 0.1 bpp.

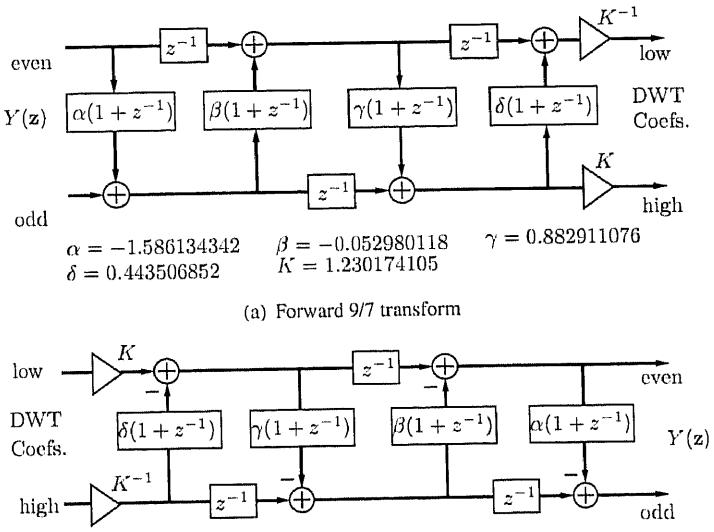


Figure 2.9: Lifting implementation of DWT in JPEG2000 of lossy compression.

for the quality recovery after decoding at high bitrates. However, the original weight is not necessarily optimal for controlling the R-D performance because the synthesis process includes the reinterlacer and the synthesis filter coefficients virtually change. In order to optimize the quality recovery, the energy gain factors have to be recalculated by taking the reinterlacer into account.

2.3.2 Discrete Wavelet Transforms in JPEG2000

Let us now briefly discuss the lifting implementation of wavelet transforms employed in JP2. Two types of wavelet transforms are standardized in JP2 Part 1 [25, 28]. One is the reversible 5/3 transform of lossless compression and the other is the irreversible 9/7 transform of lossy compression. These transforms are implemented through lifting structures composed of two-tap lifting filters. Figure 2.9 illustrates the structure of the forward 9/7 transform and its inverse. As will be shown, the invertible deinterlacer in Fig. 2.7 is followed by the forward transform and the reinterlacer follows the inverse. In this chapter, the irreversible 9/7 transform of the lossy compression only is considered, where a similar situation holds for any other lifting DWT including the reversible 5/3 transform.

Note that, for the lossless compression mode in JP2, the rounding operations are inserted after each lifting step to make transform coefficients integers [28]. As for the invertible deinterlacer shown in Fig. 2.7 (a), however, a scaling process with θ on the bottom line is required and it brings fractional bits to the transform coefficients. As a result, the application of the deinterlacers to the lossless mode requires some other special treatment for the fractional bits. This issue is investigated in Chapter 3.

Table 2.1: Synthesis filter Coefs. Integrated with Reinterlacer for $\theta = 0.5$.

n	$f_0(n)$	$f_1(n)$
0	1.11508705384879	1.47276214857289
± 1	0.65377176960586	-0.26686411943681
± 2	-0.05754352557561	-0.03144653669661
± 3	-0.15377176136265	0.01686411824750
± 4		0.04506545507608

2.3.3 Derivation of DWT Integrated with Invertible Deinterlacer

In the article [19], it was demonstrated that it is possible to integrate the lifting DWT with the (3+1)-tap deinterlacer and the inverse with the reinterlacer, where we were concerned only with the computational issues.

All of prediction and update filters of the 9/7 wavelet transform, are simple Haar-type two-tap filters with some different scaling factors (Fig. 2.9). As well, the (3+1)-tap deinterlacing filter is also composed of a Haar-type two-tap filter for the vertical direction (Fig. 2.7). Hence, the first lifting stage of the vertical wavelet transform can be combined with the deinterlacer. This chapter derives the compensated gain factor by using this fact.

The resulting polyphase matrix of the forward DWT integrated with the deinterlacer is given in Eq. 2.3, and that of the inverse transform is given in Eq. 2.4. Eqs. 2.3 and 2.4 are expressed as a Type-I polyphase matrix and a Type-II polyphase matrix, respectively [4]. From Eq. 2.3, one can verify that the first prediction of the wavelet transform and the deinterlacer can be combined. A similar situation holds for the inverse wavelet transform and the reinterlacer. Consequently, the synthesis filters are obtained as follows [4]:

$$\begin{pmatrix} F_0(z_V) & F_1(z_V) \end{pmatrix} = \begin{pmatrix} z_V^{-1} & 1 \end{pmatrix} \mathbf{R}(z_V^2). \tag{2.5}$$

Table 2.1 lists the synthesis filter coefficients obtained from Eqs. 2.4 and 2.5 for $\theta = 0.5$. The recalculated synthesis filter coefficients are used only for the level-1 vertical wavelet transform. Additionally, Table 2.2 lists the original and the compensated energy gains for the level-1 subbands, where the gain of each subband is normalized by each 1LL's gain. In each subband, the relevant gain as energy weight is just the squared norm of low- and high-pass synthesis filter coefficients. For a deeper decomposition stage, the original gain is applied in the case of multilevel DWTs.

The proposed gain compensation relies only on the MSE measure and is calculated only from the L2 norm, i.e. the squared energy, of the synthesis filters. Unfortunately, MSE is known to be poor model for the perceptual significance of the distortion. For overcoming this problem, one can apply the so-called Contrast Sensitivity Function (CSF) to frequency weighting in JP2, provided a final viewing condition [25, Annex J.12] [27, Annex B]. Note that CSF weights can additionally be applied to our derived gain factors for further subjective optimization [25, Annex J.14.4.1], although the following experimental results are obtained without applying CSF weights.

Table 2.2: Synthesis energy gains of level 1 subbands.

	Original for $\theta = 1.0$	Proposal for $\theta = 0.75$	Proposal for $\theta = 0.5$	Proposal for $\theta = 0.25$
LL	1.00000000	1.00000000	1.00000000	1.00000000
HL	0.51441208	0.51441208	0.51441208	0.51441208
LH	0.51441208	0.68510677	1.03782740	2.00093023
HH	0.26461979	0.35242720	0.53387095	1.02930268

Table 2.3: Properties of test video sequences.

Sequence	Format	Characteristics
Football	486×720i @ 60Hz 240 frames	Fast camera panning, fast moving objects
Susie	486×720i @ 60Hz 240 frames	Fixed camera view, a large moving object
Tempete	486×720i @ 60Hz 240 frames	Slow zooming out, lots of tiny moving objects
Canoa Valsesia	576×720i @ 50Hz 220 frames	Fast camera panning, a fast moving object
Rugby	576×720i @ 50Hz 220 frames	Fast camera panning, fast moving objects
Mobile & Calendar	576×720i @ 50Hz 220 frames	Slow moving camera and objects, abundant detail

Table 2.4: Average PSNR [dB] of frames decoded at middle- and high- bitrates. For deinterlacing $\theta = 0.5$ was used.

Football	0.5 bpp	1.0 bpp	2.0 bpp
Field-based	36.04	40.45	45.86
Interleaving	36.13	40.66	45.95
After Reint. (proposal)	35.57	40.18	45.34
After Reint. (default)	34.76	39.03	43.22
Before Reint. (proposal)	29.84	30.44	30.70
Before Reint. (default)	29.77	30.39	30.66
Mobile & Calendar	0.5 bpp	1.0 bpp	2.0 bpp
Field-based	24.09	28.59	36.24
Interleaving	25.80	30.44	38.04
After Reint. (proposal)	25.45	29.93	37.58
After Reint. (default)	24.77	29.16	36.32
Before Reint. (proposal)	25.06	27.93	30.02
Before Reint. (default)	24.74	27.71	29.99

2.4 Performance Evaluation

To see the improvements of the proposed method numerically, let us evaluate the coding efficiency and filter performance by selecting several values of parameter θ. The results are also compared with those obtained by the conventional field interleaving ($\theta = 1.0$) and the field-based coding. In this evaluation, several video test sequences are used. The characteristics of these sequences are summarized in Table 2.3.

2.4.1 Coding Efficiency

To demonstrate the improvement through the derived gain compensation, let us evaluate the coding efficiency in terms of PSNR for luminance components. Here, the sequences listed in Table 2.3 is used. Every frame of the sequences is encoded at 2.0 bpp and then decoded at $\{0.1, 0.25, 0.5, 1.0, 2.0\}$ bpp by the lossy mode of the JP2 Part 1 codec (JJ2000 ver.4.1 [23,26]). Figure 2.10, as objective performance evaluation, shows the R-D curves of the conventional field interleaving, the invertible deinterlacing with default gain and the deinterlacing with the compensated gain. Table 2.4 numerically shows the average PSNR of sequences "*Football*" and "*Mobile & Calendar*" when decoded at 0.5, 1.0, and 2.0 bpp, where the parameter is chosen as $\theta = 0.5$. For reference, the result of the field-based coding is also given. In addition, two examples of decoded and reinterlaced pictures of sequence "*Mobile & Calendar*" are shown in Fig. 2.11, where the default and compensated gain factors are applied for $\theta = 0.5$. Furthermore, their magnified views are shown in Fig. 2.12 as well as those of the original picture, pictures obtained by the field-based coding and the frame-based coding with field interleaving.

First of all, the quality recovery performance of reinterlacing is discussed. For sequences "*Canoa Valsesia*", "*Rugby*" and "*Football*", including fast moving objects with fast camera action, one can verify that the quality recovery with the compensated gain at higher bitrates than 1.0 bpp achieves the improvement of more than 8 dB in PSNR from the result before reinterlacing, while the results with the default gain shows worse recovery. As well, for sequences "*Mobile & Calendar*", "*Susie*" and "*Tempete*"

including slow or small moving objects with little camera action, the quality with the compensated gain is recovered more than 7 dB from the result before reinterlacing at 2.0 bpp, and is also superior to the result obtained with the default gain. In summary, compared with the default gain, the proposal always improves the quality more than 0.5 dB in decoding at more than 1.0 bpp for all test sequences as shown in Fig. 2.10, where sequences "*Canoa Valsesia*", "*Rugby*" and "*Football*" show the PSNR improvement more than 1.0 dB. From Fig. 2.12 (d) and (e), it is perceptually verified that the gain compensation recovered the picture better than the default gain.

Next, let us discuss a comparison among the field-based coding, the frame-based coding with the field interleaving and the proposal. In the experimental results shown in Figs. 2.10 (a) - (c), the sequences, which include fast camera panning as well as fast moving objects, show that both of the field-based coding and the frame-based coding with the simple field interleaving has better PSNR gain around 0.5 dB than the proposed method. In contrast, the results in Figs. 2.10 (d) - (f), for sequences including inactive camera motion with slow or tiny moving objects, show that the proposal achieved PSNR gains in the range of 0.5 - 1.5 dB relative to the field-based coding over the bitrates range of 0.1 - 2.0 bpp.

From the above experiments, it is observed that the proposed method for $\theta = 0.5$ behaves similarly to the conventional frame-based mode with field interleaving and shows performance stable to change in the characteristics of a given sequence at the price of around 0.5 dB loss in PSNR. Remember that, although the field interleaving suffers from the comb-tooth artifacts, the proposed method can suppress those disagreeable effects in the same frame-based coding mode. The detailed discussion of the deinterlacing filter performance is made in the following subsection.

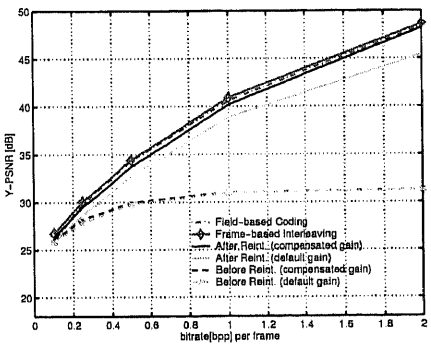
2.4.2 Filter Performance

To demonstrate the effect of deinterlacing filters with the proposed gain compensation, let us evaluate both of the comb-tooth artifact suppression capability and the coding efficiency for several choices of parameter θ.

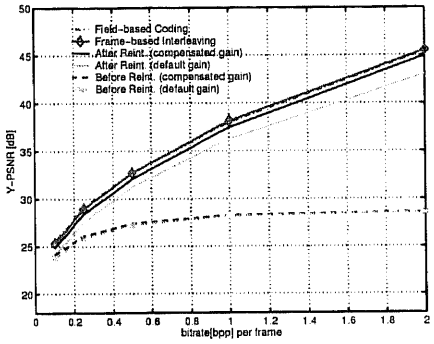
Performance of Comb-Tooth Suppression

To see how the parameter value influences to the comb-tooth suppression capability, let us evaluate some decoding results with several deinterlacing filters by selecting θ among the value $1.0, 0.75, 0.5$, and 0.25. In the first experiment, a frame of successive two fields in sequence "*Football*" is used for showing the deinterlacing effect as the case of pictures including fast moving objects with fast camera action. The frame is encoded at 2.0 bpp and then decoded at 0.25 bpp.

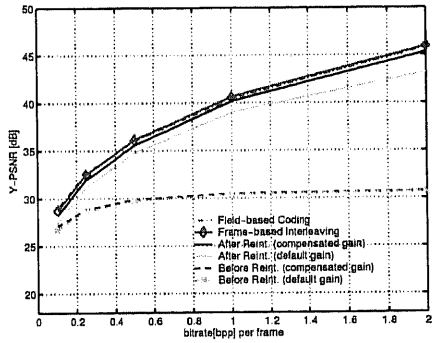
Figure 2.13 shows the whole decoded frame pictures for $\theta = 1.0$ and 0.25 and Fig. 2.14 shows the magnified views of the decoded pictures at 0.25 bpp obtained by using the field-based mode and the frame-based mode with the deinterlacing filters, where (a) shows, for reference, the result of the field-based mode and (b) - (e) show the decoded frame pictures of the frame-based mode with deinterlacing filter of θ in $\{ 1.0, 0.75, 0.5, 0.25 \}$, respectively. From Fig. 2.13, one can verify that the capability of the comb-tooth suppression around the moving objects is still maintained by the proposed gain compensation with θ smaller than one. It is also observed from Fig. 2.14 that, as the deinterlacing filter characteristic reaches to the temporal lowpass one ($\theta = 1.0$), the capability to remove the vertical high frequencies decreases so that comb-tooth artifacts remain more. In contrast, the vertical lowpass filter used in Fig. 2.14 (e)



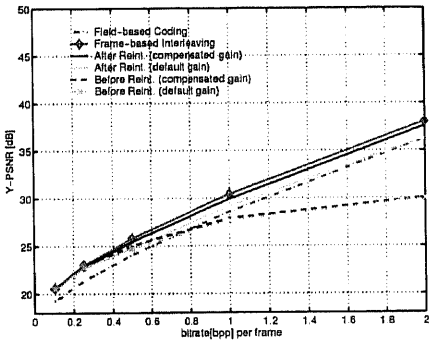
(a) *Canoa Valsesia* (576 × 720) luminance



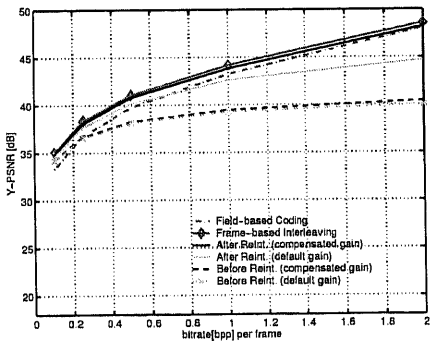
(b) *Rugby* (576 × 720) luminance



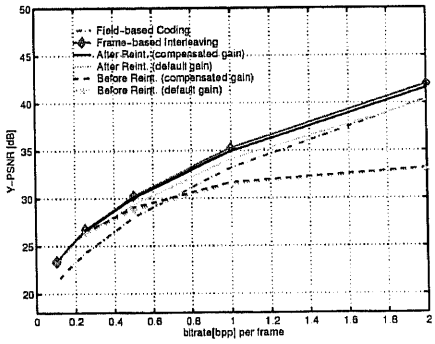
(c) *Football* (486 × 720) luminance



(d) *Mobile & Calendar* (576 × 720) luminance

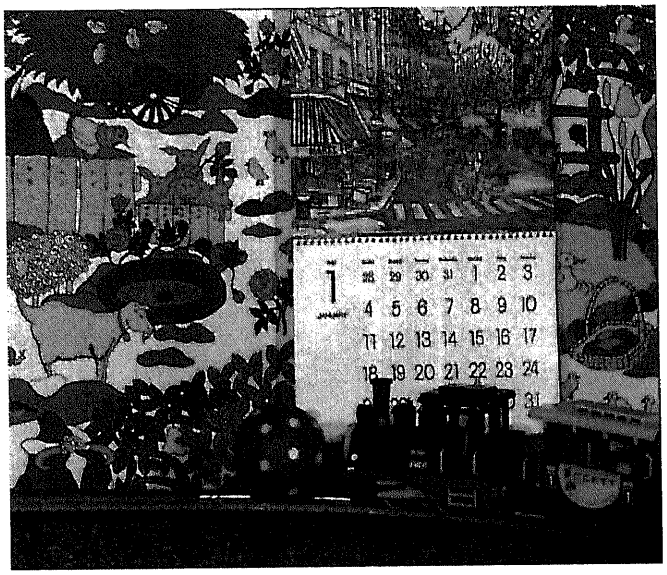


(e) *Susie* (486 × 720) luminance

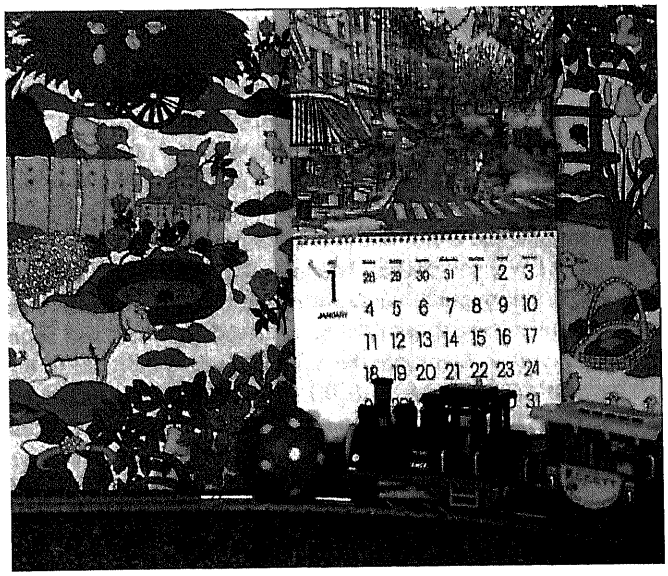


(f) *Tempete* (486 × 720) luminance

Figure 2.10: Coding efficiency for $\theta = 0.5$, where the R-D curves are obtained by taking the average of PSNR only for the luminance component (Y) over whole frames of each sequence.

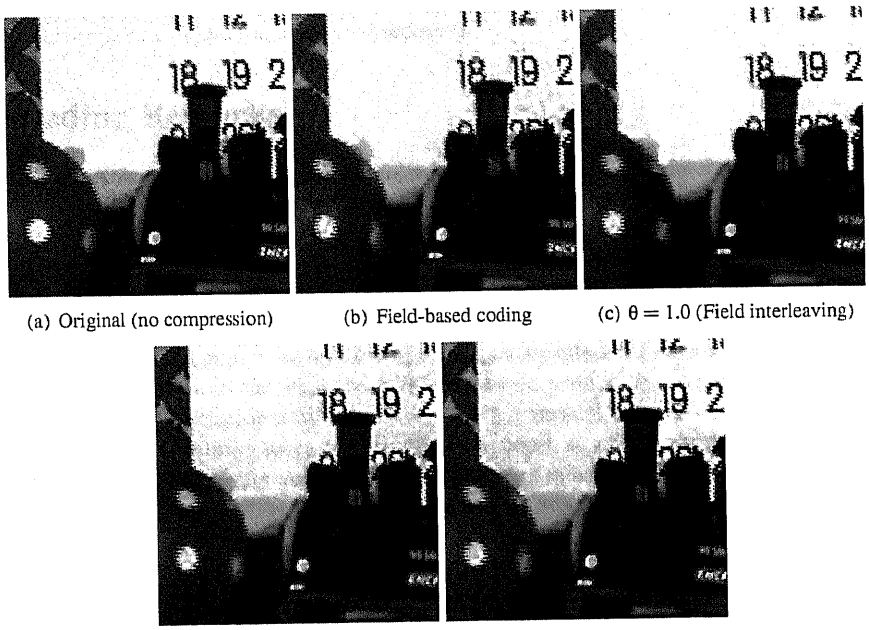


(a) After reinterlacing, where $\theta = 0.5$ (default gain)



(b) After reinterlacing, where $\theta = 0.5$ (compensated gain)

Figure 2.11: Decoded pictures at 1.0 bpp for *Mobile & Calendar*.



(a) Original (no compression) (b) Field-based coding (c) $\theta = 1.0$ (Field interleaving)

(d) After reinterlacing, where $\theta = 0.5$ (default gain) (e) After reinterlacing, where $\theta = 0.5$ (compensated gain)

Figure 2.12: Magnified views of decoded pictures at 1.0 bpp for *Mobile & Calendar*.

for $\theta = 0.25$ significantly removes the comb-tooth artifacts compared with the others.

To evaluate the deinterlacing effect for a sequence including slow moving objects with slow camera action, the second experiment shows some decoded results of a frame of successive two fields in sequence "*Mobile & Calendar*" for different values of parameter θ, where the frame is encoded at 2.0 bpp and then decoded at 0.25 bpp. Figure 2.15 shows the whole decoded pictures for $\theta = 1.0$ and 0.25 decoded at 0.25 bpp and Fig. 2.16 shows the magnified views of the results obtained by using the field-based mode and the frame-based mode with the deinterlacing filters, where (a) shows the result of the field-based mode for reference and pictures in (b) - (e) show the results of the proposed method for $\theta = 1.0, 0.75, 0.5$ and 0.25, respectively. One can verify that the comb-tooth artifacts causing around and inside of the rolling ball are significantly suppressed for a small θ.

Coding Efficiency

Lastly, several R-D curves are given in terms of PSNR in Fig. 2.17 for evaluating the coding efficiency of different values of parameter θ.

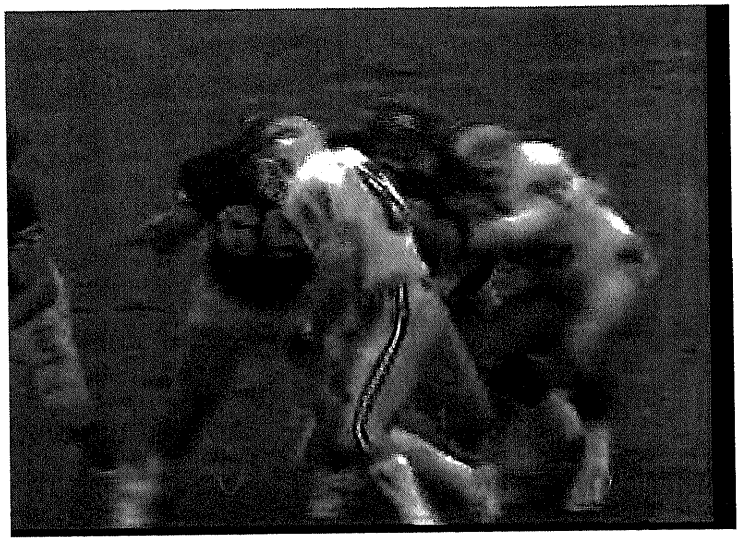
Here, sequences "*Canoa Valsesia*", "*Mobile & Calendar*", "*Football*" and "*Tempete*" are used. Every frame is encoded at 2.0 bpp, and then decoded at 0.1, 0.25, 0.5, 1.0, and 2.0 bpp. Although a filter of a small θ suppresses comb-tooth artifacts, PSNR of decoded frames is inferior to that of a filter with a large θ. This is because, as θ approaches to zero, the reinterlacing filter coefficients become infinity and the recovery performance becomes sensitive to errors. Therefore, it is found a trade-off between the

comb-tooth-artifact suppression and the quality of recovery.

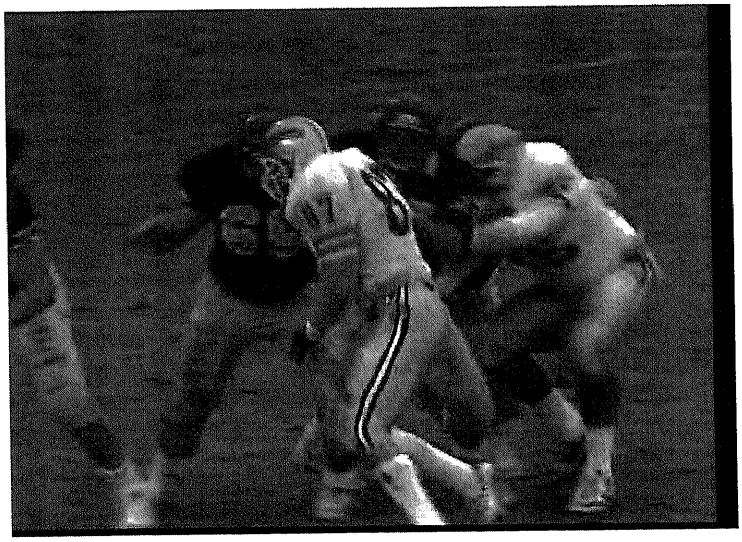
2.5 Concluding Remarks

A Motion-JPEG2000 codec system with invertible deinterlacing was presented in this chapter. To improve the quality of recovery, an energy gain compensation technique was proposed. It was verified that the quality recovery can be improved through the compensation for the energy gain factors of subbands by taking the influence of reinterlacing into account. The proposed technique has maintained the SNR scalability of the frame-based MJP2 coding system for interlaced scanning videos and shown better quality recovery than the default gain, which still keeps the filtering effect for suppressing the comb-tooth artifacts within the standard MJP2 bitstream format. It was also found that, as a (3+1)-tap deinterlacing filter becomes close to a vertical filter ($\theta \rightarrow 0$), the comb-tooth suppression capability increases. On the other hand, as a filter becomes close to a temporal filter ($\theta \rightarrow 1.0$), the quality recovery is gained much more.

In next chapter, a lossless mode of the deinterlacing technique integrated to Motion-JPEG2000 will be presented to realize a reversible frame-based coding system equipped with a similar scalable functionality for interlaced scanning videos. Currently, there are a few problems, for instance, the bit increasing due to the scaling process with θ for odd (bottom) line shown in Fig. 2.7 (a).



(a) $\theta = 1.0$ (Field interleaving)



(b) $\theta = 0.25$

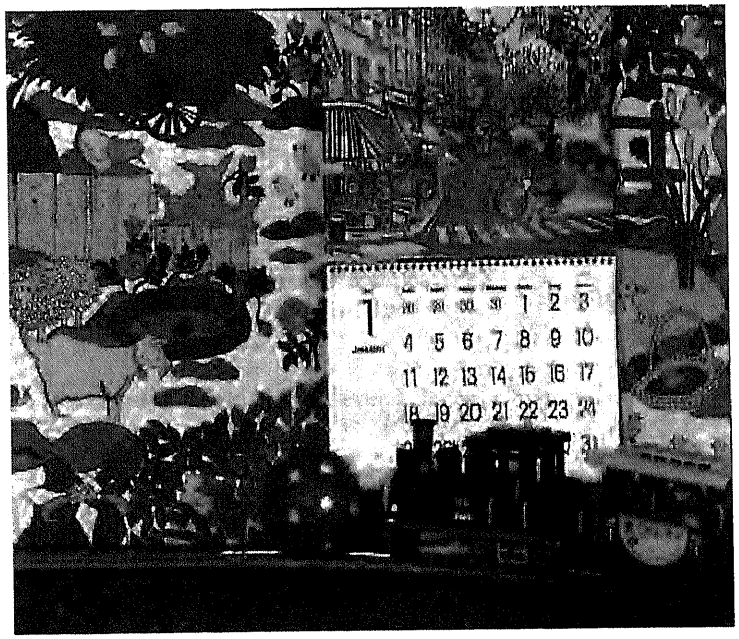
Figure 2.13: Simulation results with the gain compensation decoded at 0.25 bpp.

(a) Field-based coding (b) $\theta = 1.0$ (Field interleaving) (c) $\theta = 0.75$

(d) $\theta = 0.5$ (e) $\theta = 0.25$

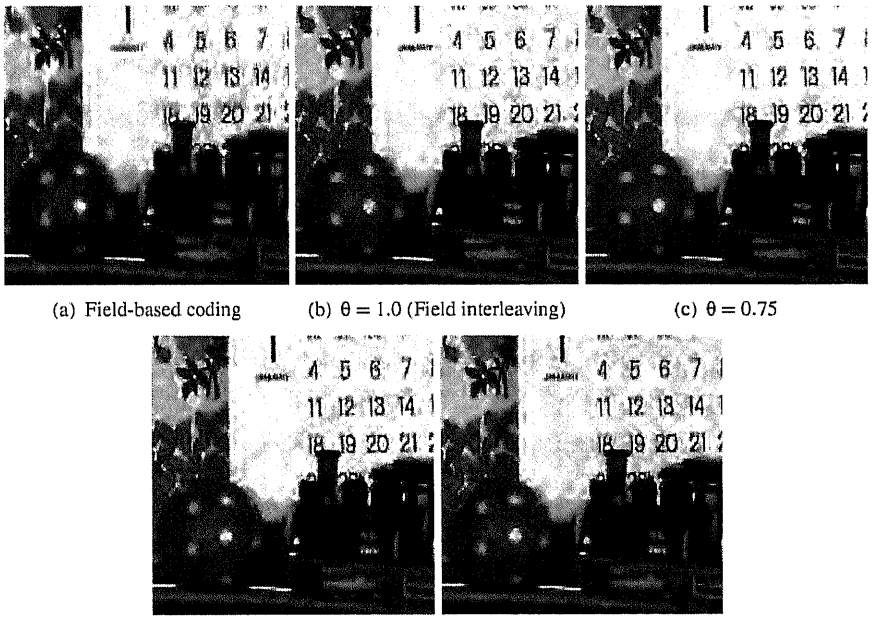
Figure 2.14: Magnified views decoded at 0.25 bpp for several choices of parameter θ.

(a) $\theta = 1.0$ (Field interleaving)



(b) $\theta = 0.25$

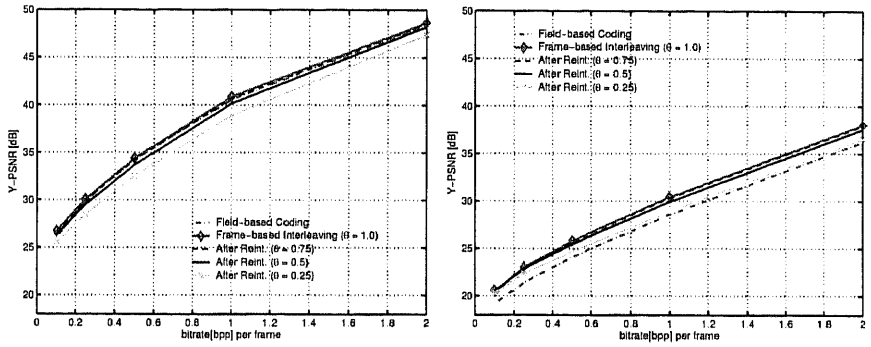
Figure 2.15: Simulation results with the gain compensation decoded at 0.25 bpp.



(a) Field-based coding (b) $\theta = 1.0$ (Field interleaving) (c) $\theta = 0.75$

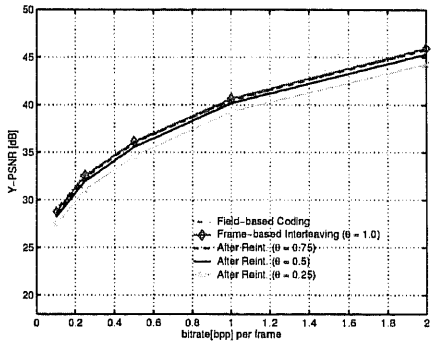
(d) $\theta = 0.5$ (e) $\theta = 0.25$

Figure 2.16: Magnified views decoded at 0.25 bpp for several choices of parameter θ.

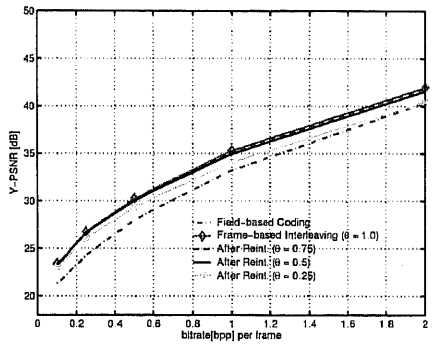


(a) R-D curve (*Canoa Valsesia*)

(b) R-D curve (*Mobile & Calendar*)



(c) R-D curve (*Football*)



(d) R-D curve (*Tempete*)

Figure 2.17: Coding efficiency for several choices of parameter θ, where the R-D curves are obtained by taking the average of PSNR only for the luminance component (Y) over whole frames of each sequence.

Chapter 3

Lossless Compression for Interlaced Video Sequences

A lossless implementation technique of Motion-JPEG2000 (MJP2) with invertible deinterlacing is presented in this chapter. In previous chapter, the invertible deinterlacing is defined as format conversion of interlaced pictures to progressive pictures of half the temporal sampling rate to the interlaced pictures, and vice verse, under the constraints of the same sampling density. This offers a reversible format conversion. In this chapter, an exact lossless implementation technique is presented for a video coding system with this technique. The reversible system can be achieved by integrating the deinterlacer into MJP2 codecs and modifying the header information for appropriate standard decoding. Some simulation results are demonstrated. It is also shown that the comb-tooth suppression capability can be kept at low bitrates with standard MJP2 decoders.

3.1 Introduction

The objective of this chapter is to deal with a reversible intra-coding system with invertible deinterlacers. The system enables the video format conversion between progressive and interlaced scanning formats without loss of information contents. Interlaced scanning and progressive scanning formats are in use for recording and displaying motion pictures [1,2,6,7]. On the other hand, one of representative intra-coding standards is Motion-JPEG2000 (MJP2) as Part 3 of the JPEG2000 (JP2) image coding standard as motion picture coding. No motion compensation is employed, and each frame is encoded independently so that handling of each frame is quite easy and errors in a frame never propagate to the others. JP2/MJP2 excellently covers the SNR, spatial, frequency, and component dominant (color) scalability. In particular, this technique is realized efficient integrated lossy/lossless coding. The SNR scalability enables us to obtain a video tailored to arbitrarily specified bit-rate with moderate quality by partially decoding a bit stream which has been encoded at higher bit-rate. This functionality is quite useful when viewers want to control the balance between quantity and quality.

For interlaced scanning videos, any intraframe-based coding requires a field interleaving (or field-merge) process so as to generate a frame picture from two successive fields. This process, unfortunately, causes horizontal comb-tooth artifacts at edges of moving objects [10]. In other words, frames generated through field interleaving suf-

fer from a tricky problem because of vertical-temporal aliasing or a vertical-temporal offset between two successive fields in the same frame [2, 6]. In the case of transform-based coding such as MJP2, errors caused by quantizing the vertical high-frequency components of comb-tooth artifacts become annoyingly perceivable especially at low bitrates. This fact implies that some remedy for the comb-tooth artifacts is required for making the most of the SNR scalability in MJP2 for interlaced scanning videos.

To solve the problem, an invertible deinterlacing technique with sampling density preservation is proposed as a preprocess for scalable intraframe-based coding [12, 13]. This technique can suppress comb-tooth artifacts at low bitrates while guaranteeing the quality recovery through the inverse process as post filtering at high bitrates [14–18]. The invertible deinterlacer was used to enhance the scalability of MJP2 effectively. In Chapter 2, a lossy coding system integrated with the invertible deinterlacer was proposed. The system suppresses comb-tooth region in frame pictures generated from interlaced pictures and is applicable to both of frame and field based-display applications with single code-stream. It is easily imagined that those frame pictures are frequently edited and recompressed regardless of being used in professional or consumer applications. Usually, it is required to preserve all accurate information during such repetitive processes. The reversible, i.e. lossless, conversion supporting to reconstruct the interlaced images from the deinterlaced code-stream is therefore of interest. Lossless compression is desirable in applications where the images are to be extensively edited and recompressed, whereas the accumulation of errors by multiple lossy compression might become unacceptable.

In this chapter, the lossless mode of MJP2 is presented taking such a reversible deinterlacing technique into account. To achieve this purpose, a MJP2 codec system integrating with lossless deinterlacing is discussed subject to the standard code-stream format and proper low bit-rate standard decoding. Since the invertible deinterlacing technique can be implemented by a lifting structure for a special case, it is easy to be integrated into the DWTs of MJP2 [19]. It is shown that the lossless compression for interlaced scanning video sequences is achieved by controlling the subband gain factors [28] according to the DWT integrated with the invertible deinterlacing. In addition, the proposed method offers a function of decoding a picture of which comb-tooth artifacts have already been suppressed. This function is realized within the standard code-stream by modifying a component of header marker in JPEG2000 (JP2). The performances of the proposed method are evaluated, and the simulation results show the significance compared with normal field interleaving.

3.2 Lifting Implementation of Invertible Deinterlacer

In a special case of deinterlacing and reinterlacing filters [12, 13], the implementation is available in a lifting structure. Figure 3.1 shows the corresponding lifting implementation, where z_V^{-1} indicates the delay for vertical direction. Each structure is composed of one prediction with the simple Haar type 2-tap filter and the scaling factor on the odd line. These lifting structures are equivalently applied to the vertical direction of interleaved frame pictures instead of directly handling interlaced scanning video sequences. Those filters have the following properties: i) normalized amplitude to keep brightness, ii) regularity to avoid the checkerboard effect [20, 21], and iii) vertical symmetry to afford the symmetric border extension [20, 22]. It is available to select any value of the design parameter θ among the range $0 < \theta \leq 1$ for the pair shown in Fig. 3.1. The chapter exemplifies four frequency characteristics of the deinterlacing filter $H(\mathbf{z})$ in

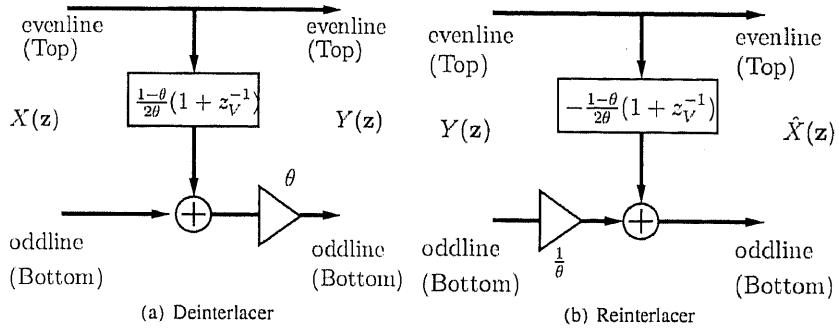


Figure 3.1: Lifting implementation.

Fig 3.2, where ω_V and ω_T indicate the normalized angular frequencies for vertical and temporal direction, respectively. The characteristics are changed between temporal, vertical-temporal (V-T), vertical lowpass filters by selecting θ.

In JP2 part I [28], two types of wavelet transforms are standardized. One is the reversible 5/3 transform of lossless compression and the other is the irreversible 9/7 transform of lossy compression. These transforms are implemented as lifting structures with two-tap filtering. Additionally, the invertible deinterlacer can be cascaded to the following wavelet transform. Figure 3.3 illustrates the lifting structures where deinterlacing and reinterlacing are embedded into the 5/3 transforms, respectively.

3.3 Lossless Implementation

In this section, we present a lossless compression technique of MJP2 integrated with the invertible deinterlacer. There are the following two issues to realize the lossless compression:

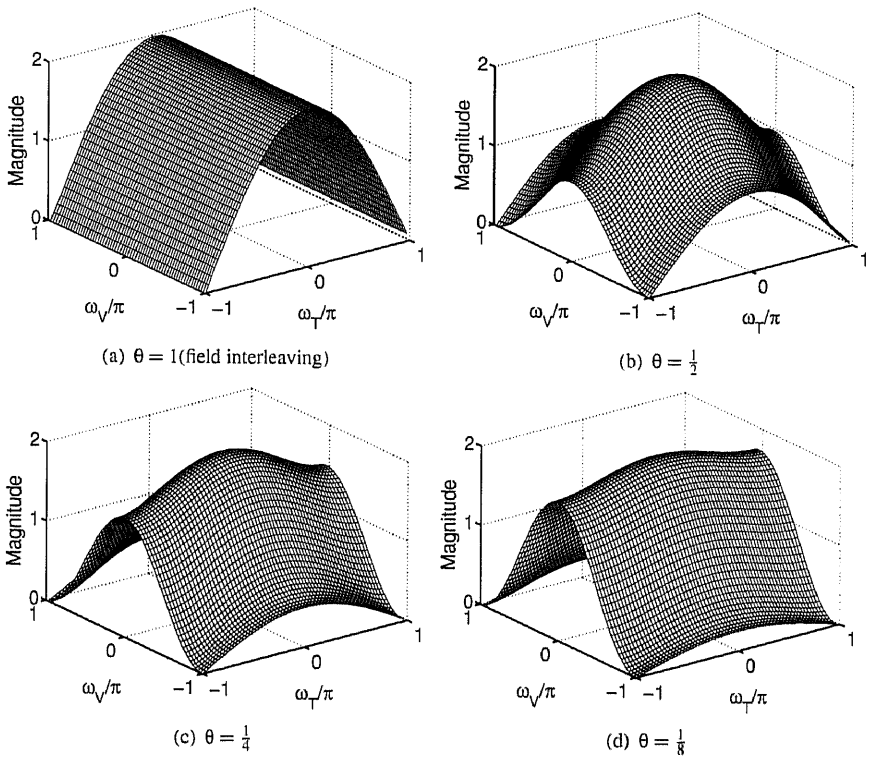
- How to prevent high frequency subband's fractions due to output scaling with factor θ
- How to maintain standard decoding and the capability of comb-tooth suppression

3.3.1 Preservation of fractions

As mentioned, the 5/3 transform is employed for lossless compression in JP2. In order to realize the reversible wavelet transform, the rounding operations are inserted after the lifting steps of '*prediction*' and '*update*' to make transform coefficients integers [28]. The bit-depth R_b of wavelet coefficients of subband b is represented by the nominal range R_I of the input subband and the subband gain factor G_b as follows:

$$R_b = R_I + \log_2(G_b) + GB, \tag{3.1}$$

where G_b is given by the lowpass and high-pass filters of wavelet transform as shown in Table 3.1, and GB denotes guard bits. For the lifting structure integrated with invertible deinterlacing shown in Fig. 3.3 (a), the process of scaling on odd line generates fraction parts. As a result, the lossless mode becomes unavailable. In this study, the gain of the



(a) $\theta = 1$(field interleaving)

(b) $\theta = \frac{1}{2}$

(c) $\theta = \frac{1}{4}$

(d) $\theta = \frac{1}{8}$

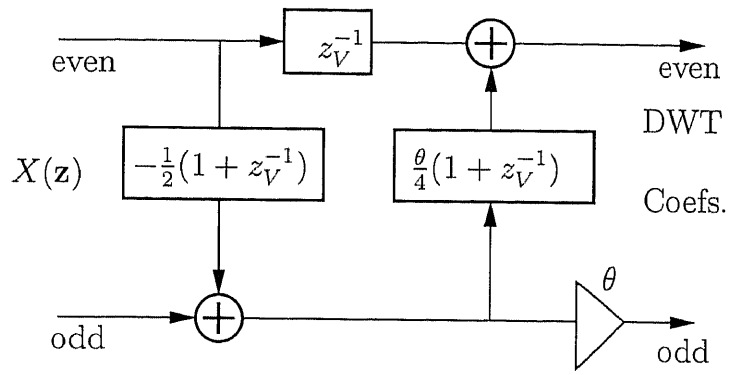
Figure 3.2: Characteristics of $H(\mathbf{z})$, where only the vertical-temporal frequency plain is shown. ω_V and ω_T indicate the normalized angular frequencies for vertical and temporal direction, respectively.

high-pass filter is obtained by taking the scaling factor θ into account to preserve the fractions.

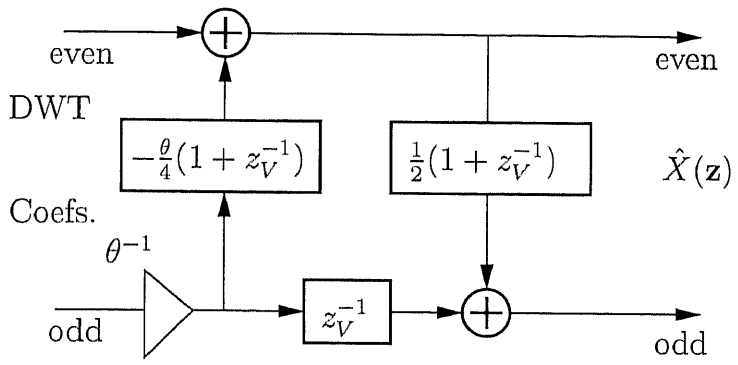
Table 3.2 shows the derived subband gain factors. According to the result, modifying the bit-depth of wavelet coefficients guards the fractions that would have been eliminated in the original gains. Note that parameter θ can be given by 2^{-n} with a positive integer n and that the proposed gains are only applied for the first vertical wavelet decomposition ($b = 1$).

3.3.2 Maintenance of standard decoding

The important issue of the lossless implementation is how to guarantee the appropriate decoding with the standard code-stream format and maintain the suppression capability of comb-tooth artifacts at low bitrates. Modification of the subband gains makes the high frequency components larger than the original bit-depth range defined in JP2. To preserve the fractional bits, it is proposed to handle them with the entropy coding process through gain modification instead of discarding them. Introducing the DWT integrated with the deinterlacer without the scaling factor θ means to raise the high-pass DWT coefficients. As a result, the dynamic ranges of LH_1 and HH_1 subbands are increased by $\theta^{-1}(> 1)$ times as large as the original. When using the inverse DWT



(a) 5/3 forward transform



(b) 5/3 inverse transform

Figure 3.3: Lifting implementation of the 5/3 transform integrated with the deinterlacer and its inverse.

shown in Fig. 3.3 (b) without the scaling factor θ on the odd line, the brightness is kept in the reconstructed pictures. However, with the ordinary standard decoders, the gain raising influences the quality of decoded pictures, and produces disagreeable results. Thus, the direct application of the proposed gains makes the SNR scalability of MJP2 meaningless.

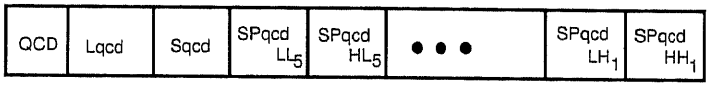
In JP2, the subband gains are written to the header information with the marker QCD (Quantization Default) [28] as shown in Fig 3.4. In this study, the original gains are forcibly written in segments of the marker QCD in the code-stream despite of applying the proposed gains in Table 3.2 during the encoding process. In the case of the reversible compression, the Sqcd parameters consist of the number of guard bits and a value which indicates '*no quantization*.' Fortunately, the following SPqcd parameter of each subband simply becomes the reversible dynamic range except for GB, i.e. $R_b - GB$. Thus, we simply replace the compensated gain shown in Table 3.2 with the original gain only for the SPqcd values of LH_1 and HH_1 when encoding. We write the header information by using the original gain, and the wavelet coefficients of each subband are quantized and encoded by taking the compensated gains of LH_1 and HH_1 into account. Summarizing, the encoding process steps are as follows:

Table 3.1: Original gain

Subband b	$G_b^{default}$
LL_b	1
HL_b	2
LH_b	2
HH_b	4

Table 3.2: Proposed gain

Subband 1	$G_1^{comp.}$
LL_1	1
HL_1	2
LH_1	$2\theta^{-1}$
HH_1	$4\theta^{-1}$



QCD: Marker value (0xFF5C)
Lqcd: Length of marker segment in bytes
Sqcd: Quantization style for all components
SPqcd: Quantization step size value for each subband

Figure 3.4: Quantization default syntax in the case of the decomposition level 5.

- Encoding

  i Transform with the forward 5/3 transform integrating the deinterlacer

  ii Write the default gains into the SPqcd parameters of QCD

  iii Decide the dynamic range in the quantization step by taking the gains in Table 3.2 into account

In the case of the losslessly decoding, decoders need to know the value of parameter θ. In this chapter, assuming that the parameter θ is fixed to a single value, we do not count the amount of information to be sent to decoders. Of course, it is possible to choose preferable values of the parameter θ. In this case, we can use an option of the file format in Motion-JPEG2000 [27]. In MJP2, '*box*' is defined for the specifications of files. Among several defined boxes, *User Data Box* provides a facility for informative user-data in code-streams. The box does not affect decoding process with standard decoders. The amount of information depends on how frequently the parameter θ is transmitted. The conjecture is that it is neglectable compared with picture data.

As the dequantization step of JP2 decoders, an appropriate bit-depth of the transform coefficients are obtained by using the original gains in the segments of the marker QCD. Such a stream deceives standard decoders as if itself has been normally encoded. What happens in standard decoders is that the LH_1 and HH_1 subband's least significant bits (LSB), i.e. the fraction parts are just discarded. As a result, we can guarantee standard decoding with the standard code-stream format. It only differs in the way that comb-tooth artifacts are suppressed beforehand.

For the lossless decompression with the proposed method, it is sufficient to dequantize the vertical high frequency components with the proposed subband gain which is derived by taking the scaling factor of the deinterlacing into account, and the inverse transform integrated with reinterlacing as shown in Fig. 3.3 (b). Remind that the factor θ^{-1} has not been applied for presenting their least significant bits of the coefficients on odd lines. Thus, scaling with the factor θ^{-1} on odd lines is not needed. The procedures are summarized as follows:

- Decoding with a standard decoder
  - i Decode header information and data stream
  - ii Read the SPqcd parameters of QCD
  - iii Decide the dynamic range in the dequantization step, so that the fractional parts of LH_1 and HH_1 are discarded
  - iv Transform with the inverse 5/3 transform
- Decoding with a decoder integrated with the reinterlacer
  - i Decode header information and data stream
  - ii Read the SPqcd parameters of QCD
  - iii Compensate the gain by adding $\log_2(\theta^{-1})$, which is the scaling factor of reinterlacing in Fig. 3.3 (b)
  - iv Decide the dynamic range in the dequantization step
  - v Transform with the inverse 5/3 transform integrated with the reinterlacer

3.4 Experimental Results

To verify the significance of the proposed method, let us evaluate the comb-tooth suppression capability and the coding efficiency among several values of θ and compare them with simple field interleaving ($\theta = 1$) and field-based coding. In this evaluation, several test sequences are used, including ITE standard sequence *"Whale Show"* as a high definition (HD) sequence. The characteristics are summarized in Table 3.3.

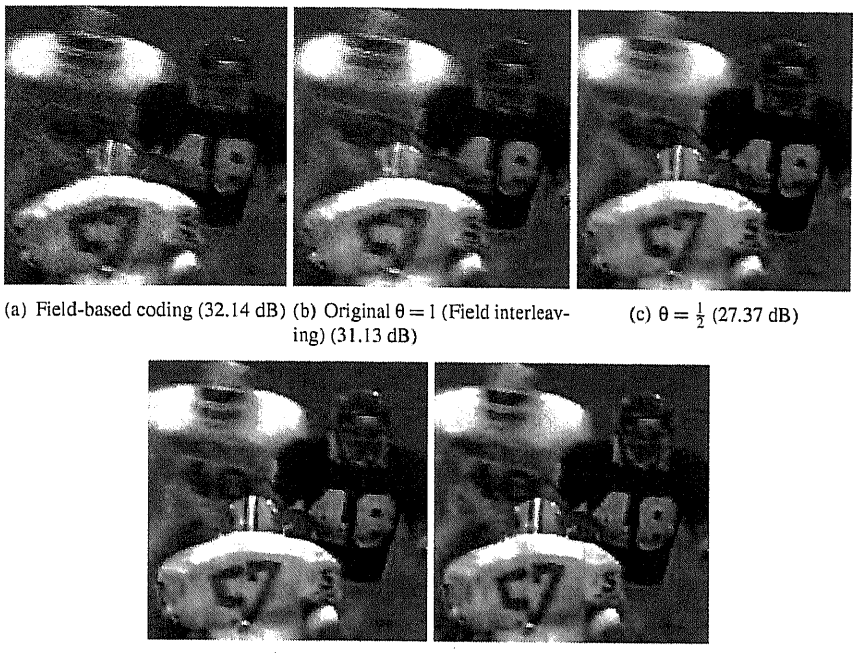
3.4.1 Suppression capability at low bitrates

To demonstrate the effect of deinterlacing filters, several decoding results are evaluated through different deinterlacing filters by selecting θ among values $1, \frac{1}{2}, \frac{1}{4}$ and $\frac{1}{8}$. In this experiment, a frame of successive two fields in sequence *"Football"* was used for showing the deinterlacing effect as the case of pictures including fast moving objects with fast camera movement. The frame was encoded with the proposed method in the lossless mode, and then decompressed with the standard JP2 decoder [23] at 0.25 bpp.

Figure 3.5 shows enlarged regions of the results for $\theta \in \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}\}$ and each PSNR. For reference, the result of field-based coding is also shown in Fig. 3.5 (a). When compared with the proposed method in terms of PSNR, the simple field interleaving shows better performance. The field-based approach outperforms these frame-based approaches in the case of *"Football"*. As seen from the pictures, however, the field interleaving suffers from the comb-tooth artifacts, which can not be recognized through PSNR. As well, the field-based approach is inferior in terms of the comb-tooth suppression capability to the proposed method despite of showing the best performance in terms of PSNR. From subjective point of view, it is verified that the proposed method can improve the comb-tooth suppression capability compared with field interleaving ($\theta = 1$) and field-based coding. The temporal lowpass filter ($\theta = 1$) shown in Fig. 3.2 (b) has less capability to remove the vertical high frequencies. As a result, comb teeth are strongly perceived at low bitrates. It is verified that the filter shown in Fig. 3.2 (b) at $\theta = \frac{1}{8}$ can significantly remove the comb-tooth artifacts.

Table 3.3: Properties of test video sequences.

Sequence	Format	Characteristics
Mobile & Calendar	576×720i @ 50Hz 220 frames	Slow moving camera and objects, abundant detail
Tempete	486×720i @ 60Hz 240 frames	Slow zooming out, lots of tiny moving objects
Susie	486×720i @ 60Hz 240 frames	Fixed camera view, a large moving object
Canoa Valsesia	576×720i @ 50Hz 220 frames	Fast camera panning, a fast moving object
Rugby	576×720i @ 50Hz 220 frames	Fast camera panning, fast moving objects
Football	486×720i @ 60Hz 240 frames	Fast camera panning, fast moving objects
Whale Show	1920×1035i @ 60Hz 450 frames	Fast camera panning and zooming, fast moving large object, abundant detail in background



(a) Field-based coding (32.14 dB) (b) Original $\theta = 1$ (Field interleaving) (31.13 dB) (c) $\theta = \frac{1}{2}$ (27.37 dB)

(d) $\theta = \frac{1}{4}$ (24.80 dB) (e) $\theta = \frac{1}{8}$ (23.71 dB)

Figure 3.5: Enlarged partial regions of decoded pictures at 0.25 bpp in *Football.*

As well, in order to evaluate the deinterlacing effect for a sequence including slow moving objects with slow camera motion, let us show some decoded results of a frame of successive two fields in sequence "*Mobile & Calendar*" for different values of parameter θ as the second experiment, where the frame was encoded in the lossless mode and then decoded at 0.25 bpp. Figure 3.6 shows the enlarged partial regions of each result and their PSNR. In contrast to "*Football,*" the field-based approach becomes worse in terms of PSNR. For the results in Figs. 3.6 (b) - (e), there is a similar tendency to "*Football.*" From subjective view point, one can recognize the degradation around these comb teeth, in particular, for the results in Figs. 3.6 (a) and (b). One can verify that the comb-tooth artifacts causing around and inside of the rolling ball were significantly suppressed with a small θ.

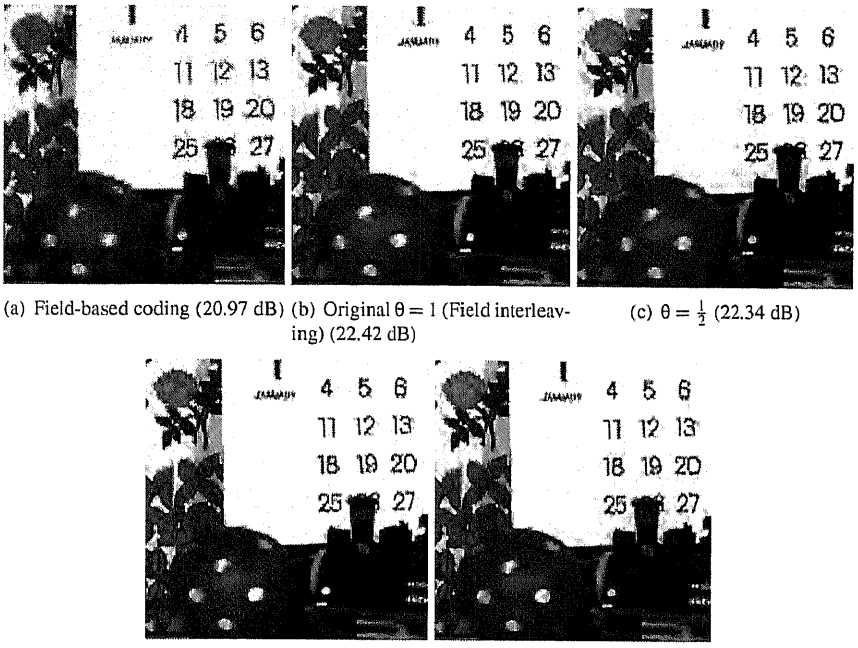
3.4.2 Coding efficiency

In order to show the significance of the proposed method, let us evaluate the coding efficiency with the lossless codec. In this experiment, the sequences listed in Table 3.3 was used. Every frame picture was encoded in the lossless mode. Table 3.4 indicates the results of the average bit-rate in terms of luminance component for every sequence.

For sequences "*Canoa Valsesia,*" "*Rugby,*" "*Football,*" and "*Whale Show,*" including fast moving objects with fast camera motion, one can verify a few bit-rate increases with any θ compared with frame-based coding ($\theta = 1$) although the deinterlacing effect strongly suppresses the artifacts at low bitrates. The rate increasing is in the range between 0.9 % to 2.3 % among these sequences. The coding efficiency of the proposed

Table 3.4: Coding efficiency (average amounts of each sequence) in lossless mode.

Sequence	Field-based coding	Frame-based coding	Proposed ($\theta = 2^{-1}$)	Proposed ($\theta = 2^{-2}$)	Proposed ($\theta = 2^{-3}$)
Canoa Valsesia	3.8599 [bpp]	3.8346 [bpp]	3.8755 [bpp]	3.9077 [bpp]	3.9216 [bpp]
Rugby	4.3246 [bpp]	4.3213 [bpp]	4.3613 [bpp]	4.3930 [bpp]	4.4086 [bpp]
Football	4.1408 [bpp]	4.1288 [bpp]	4.1658 [bpp]	4.1988 [bpp]	4.2143 [bpp]
Whale Show	4.9313 [bpp]	4.8555 [bpp]	4.8818 [bpp]	4.9088 [bpp]	4.9229 [bpp]
Mobile & Calendar	6.0750 [bpp]	5.7779 [bpp]	5.7867 [bpp]	5.8006 [bpp]	5.8067 [bpp]
Tempete	5.1849 [bpp]	4.9204 [bpp]	4.9280 [bpp]	4.9375 [bpp]	4.9401 [bpp]
Susie	3.7487 [bpp]	3.6050 [bpp]	3.6275 [bpp]	3.6416 [bpp]	3.6443 [bpp]



(a) Field-based coding (20.97 dB) (b) Original $\theta = 1$ (Field interleaving) (22.42 dB) (c) $\theta = \frac{1}{2}$ (22.34 dB)

(d) $\theta = \frac{1}{4}$ (22.08 dB) (e) $\theta = \frac{1}{8}$ (21.85 dB)

Figure 3.6: Enlarged partial regions of decoded pictures at 0.25 bpp in *Mobile & Calendar*.

methods are closer to that of field-based coding when compared with that of frame-based coding ($\theta = 1$). For sequence "*Whale Show*," the coding efficiency is improved compared with field-based coding in any case of the proposed method.

As well, for sequences "*Mobile & Calendar*," "*Tempete*," and "*Susie*," including slow or small moving objects with little camera action, one can verify that the results of the proposed method are almost the same as those of frame-based coding with field interleaving ($\theta = 1$). The increasing rate is only in the range between 0.2% to 1.0% among these sequences. For any case of the proposed methods, the coding efficiency is improved compared with field-based coding. Note that the proposed method gives moderate results at both of low bitrates and the lossless mode.

The smaller θ becomes, the more the fractional bits are required. The bit increase in high-pass coefficients sensitively influences the coding efficiency for the lossless coding because the proposed method handles the fraction in the entropy coding process by giving the gain instead of applying the scaling factor.

As a result, when comparing coding efficiency among the different deinterlacing filters in Table 3.4, we can see that the more fraction bits in the LH_1 and HH_1 subbands, the less the coding efficiency results in. In summary, there is a trade-off between the comb-tooth capability of suppression and coding efficiency.

3.5 Concluding Remarks

In Chapter 3, a lossless implementation technique of Motion JPEG2000 integrated with invertible deinterlacing was presented. It was shown that recalculating the subband gain factors and modifying the QCD marker segments enables us to realize a lossless compression compensated for interlaced scanning video sequences. Some experimental results showed that the coding efficiency is almost the same as that obtained by the encoder with normal field interleaving, while maintaining the suppression capability of comb-tooth artifacts through the standard decoding.

Chapter 4

Field Scalability for Scalable Video Coding

In this chapter, motion compensated three-dimensional (3-D) filter banks providing as new tools of spatio-temporal resolution control are presented. For recent developments in scalable video coding, most of them are based on 3-D wavelet transform with motion compensation. To achieve the function of frame-rate and spatial resolution scalabilities, motion compensated temporal filtering (MCTF) through lifting wavelet transform currently attracts many researchers as an effective temporal decomposition tool.

In the following sections, two types of systems providing the video format control between the progressive and interlaced scanning manner are discussed. Unlike other filter banks, the proposed systems are constructed in a way unique to multidimensional systems. This paper suggests two kinds of interlaced video formats: the vertical-temporal (VT) quincunx and face-centered-orthorhombic (FCO) format. Some experimental results show novel functions and the significance of the proposed filter banks.

4.1 Introduction

The objective of scalable video compression is to encode a video once and then partially decode the bit-stream at a lower bit-rate, spatial resolution and/or frame rate in the best possible quality as if it had been optimized to each decoding condition. Network video applications such as interactive video library and Internet video streaming require rate control over a wide range of bit-rates because there must be various kinds of terminals, and available channel bandwidth for each terminal changes largely moment by moment. Scalable video coding achieves this control simply by truncating the bit-streams after the source has been compressed. As well, in the application to remote video retrieval, users may receive the desired video contents by refining the bit-streams.

There are some implementations to approach this goal. MPEG-2 video defines traditional scalable coding, where a video is encoded into a base layer and a few enhancement layers [7, 8, 48]. The control is simply achieved by switching if enhancement layers are used or not, where the enhancement layers add spatial, temporal, and/or SNR quality to the reconstructed base layer. As a new form of scalability, fine granular scalability (FGS) has been developed and adopted by the MPEG-4 Visual standard [49], which allows a finer control of SNR quality in the enhancement layer. The

FGS technique is based on bit-plane coding and can be combined with the temporal scalability [50].

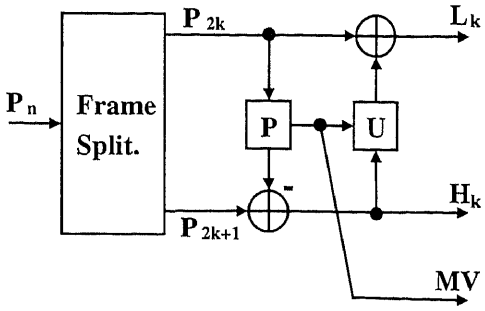
Recent developments in such video coding aim at providing fine granular control of quality, frame-rate and spatial resolution. For providing seamless adaptation with different visual quality, frame-rate and resolution to meet the requirements of end users, most of them employ 3-D wavelet-based techniques with motion compensation, in which motion compensated temporal filtering (MCTF) is preformed. For the spatio-temporal scalability, filter banks play an important role in scalable video coding. In the article [32], a three-dimensional (3-D) lifting DWT scheme with MCTF is introduced, in which a spatio-temporal transform precedes embedded quantization and bit-plane coding as done in the current state-of-the-art scalable codec for still pictures, namely JPEG2000 [28]. Currently, many researchers are working on the fine control of spatio-temporal resolution in scalable video coding through 3-D wavelet with MCTF [32–43].

In the article [51], the extension of H.264/AVC hybrid video coding towards scalable video coding (SVC) using MCTF is presented. In October 2005, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG reported the Joint Scalable Video Model JSVM-4 as an Amendment of their H.264/AVC standard [52] in which MCTF is used as an alternative and non-normative encoder configuration.

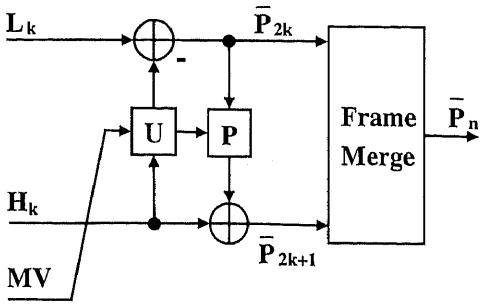
Until now, we have two choices of the spatio-temporal resolution control with respect to video formats. A system with MCTF handles a video only in the progressive scanning manner, and the other divides a video into an interlaced, or non-rectangular sampled, field sequence and its residual [54]. The former can achieve a finer scalability through multi-stage tree-structured filter banks. In contrast, the latter is not suitable for such frameworks since the subband signals are not in the same progressive manner as the input. However, the interlaced subband signals have an advantage of the preservation of update temporal interval with reduced frame rate. In particular, the existing MCTF has the following problem [51]. MCTF separately controls temporal resolution as frame rate thorough lifting wavelet transform in temporal direction. Haar, 1/3, and 5/3 transform are popularly used with MC in lifting structure for realizing MCTF. Haar and 5/3 transform are implemented by both of prediction and update steps. MCTF with update step has an advantage in the coding performance gain compared to one without update step. The update step, however, brings several issues to MCTF such as bit depth, memory band width, delay and complexity. On the other hand, 1/3 transform can be implemented very simply only with prediction step or multi-reference framework. A price is degradation in the coding efficiency. Especially, PSNR fluctuation is one of the main issues.

In this chapter, a new option to the spatio-temporal resolution control is discussed. To find a better solution, two types of systems with a spatio-temporal split process are presented: deinterlacer banks and non-separable motion compensated spatio-temporal filtering (MCSTF). Both of these techniques are novel, but simple, three dimensional filter bank and fall into the MCTF category. The proposed techniques, however, enjoy a definitely different function from others with respect to handling of video formats. The techniques can provide a new option of spatio-temporal scalability with video format control between progressive and interlaced scanning, which we refer to as *field scalability*. Two kinds of the interlaced video formats are suggested in this chapter: the vertical-temporal (VT)-quincunx and face-centered-orthorhombic (FCO) formats.

In the following sections, firstly, perfect reconstruction (PR) deinterlacer banks using invertible deinterlacers are suggested for spatio-temporal scalable video codec [44, 45]. A deinterlacer bank separates a progressive video into two different progressive videos of a half frame rate, where interlaced videos are given as intermediate data



(a) Analysis bank



(b) Synthesis bank

Figure 4.1: Lifting representation of motion compensated temporal filtering, where P, U indicate the prediction step and update step.

during the analysis process. As well, interlaced videos are also obtained as intermediate data during the synthesis process through the inverse transform.

As the alternative approach, non-separable motion-compensated spatiotemporal filtering (MCSTF) is presented. This technique enables us to provide the video format control between progressive and interlaced scanning as well as the deinterlacer banks. It is suggested to introduce the field prediction with MC techniques in order to reduce prediction error produced with the prediction step. This proposed method has different structure with MC field prediction from the deinterlacer banks. Moreover, it is shown that the proposed deinterlacer banks and MCSTF have the suppression of the PSNR fluctuation, which becomes a problem through the existing MCTF without update step. Some experimental results show the effectiveness of these proposed techniques.

4.2 Motion Compensated Temporal Filtering

This section begins by reviewing the temporal decomposition using a temporal filtering with motion compensation for video coding schemes. Since motion compensated temporal filtering is regarded as a subband/wavelet transform, the technique is available to implement a lifting scheme [46]. The lifting scheme guarantees perfect reconstruction (PR) of the input source. This property is valid even if non-linear operations are used

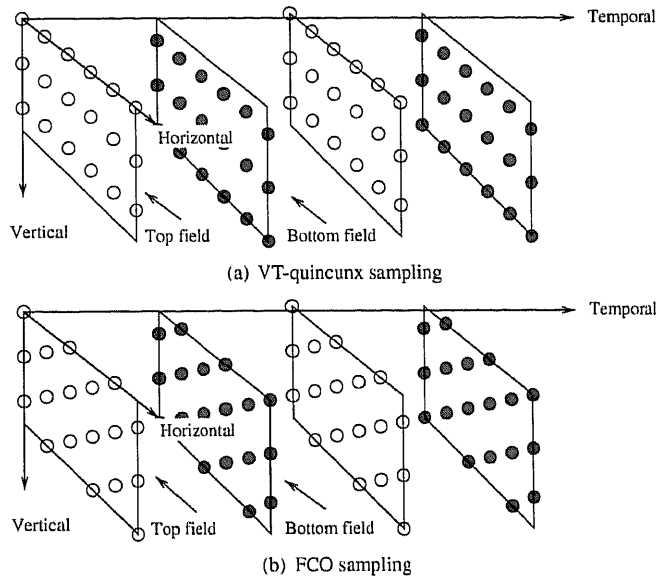


Figure 4.2: Sampling lattices, where the white and black circles are pixels on top and bottom fields, respectively.

during the lifting operation.

The lifting schemes used in video processing consist of three steps: frame splitting, prediction, and update as shown in Fig. 4.1. The input signal P_n to the analysis filter bank corresponds to a progressive video sequence. Firstly, the frame split process separates the input signal P_n to two frame sequences with even ($2k$) and odd ($2k+1$). Secondly, the frame sequence P_{2k+1} is predicted using MC and the prediction error H_k is generated. Finally, the frame sequence P_{2k} is updated and the low-pass band frame sequence L_k is given as frame sequence of a half frame-rate. MCTF with update step provides coding efficiency and reduces the PSNR fluctuation which occurs in MCTF without update step. The update step, however, involves several problems which are bit-depth memory band width, reference picture buffering, delay and complexity issues [32,47,51,53].

4.3 Perfect Reconstruction Deinterlacer Banks

For developing a novel spatio-temporal scalability, this thesis discusses two proposal techniques for scalable video coding. In this section, firstly, perfect reconstruction (PR) deinterlacer banks using invertible deinterlacers are presented.

4.3.1 Sub-Sampling Format

In this chapter, two types of sub-sampling formats are used. Figure 4.2 summarizes the video formats: the popular Vertical-Temporal (VT)-quincunx and Face-Centered-Orthorhombic (FCO) sampling formats. This chapter refers to the sampling format shown in Fig. 4.2 (a) as *line-based field sequence* and Fig. 4.2 (b) as *point-based field*

sequence for the sake of consistent presentation. In the following sections, let us refer to a progressive video format as *frame sequence* and to an interlaced video format in either of the VT-quincunx or FCO formats as *field sequence*. Since the FCO sampling fairly handles the vertical and horizontal direction, it is preferable for the human visual perception when compared with the VT-quincunx sampling.

4.3.2 Spatio-temporal Split Process

First of all, line- and point-based field split process are suggested. Figure 4.3 (a) shows the line-based case. In the same way, the point-based field split process is defined as shown in Fig. 4.3 (b). In these processes, input frame sequence P_n is divided into two field sequences Q_k and R_k, where subscripts n and k denote the frame numbers of the full and half frame rate, respectively. In the figure, $P_{\mathrm{T}n}$ and $P_{\mathrm{B}n}$ are top and bottom fields of P_n, and are associated with Q_k and R_k as $Q_k = \{P_{\mathrm{T}2k}, P_{\mathrm{B}2k+1}\}$ and $R_k = \{P_{\mathrm{T}(2k-1)}, P_{\mathrm{B}2k}\}$, respectively.

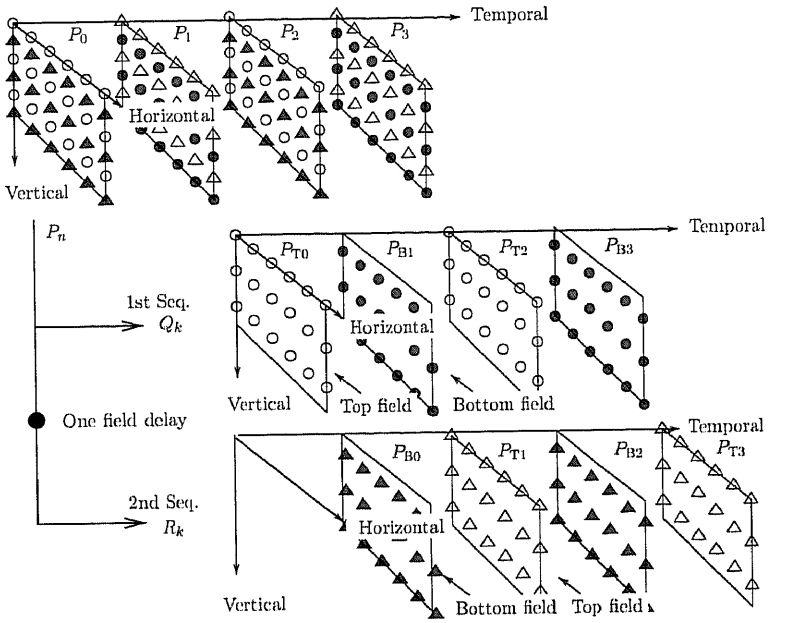
4.3.3 FCO-based Invertible Deinterlacer

In the previous chapters, an invertible deinterlacer is a kind of multidimensional sampling lattice converter, which was originally developed as a pre-process of video encoding to suppress comb-tooth artifacts caused by field interleaving for line-based interlaced videos. This technique is different from a major deinterlacer [2, 3] in that the sampling density is preserved and the original field sequence can be reconstructed by an inverse process. Let us review the process for the popular VT quincunx and newly introduced FCO format.

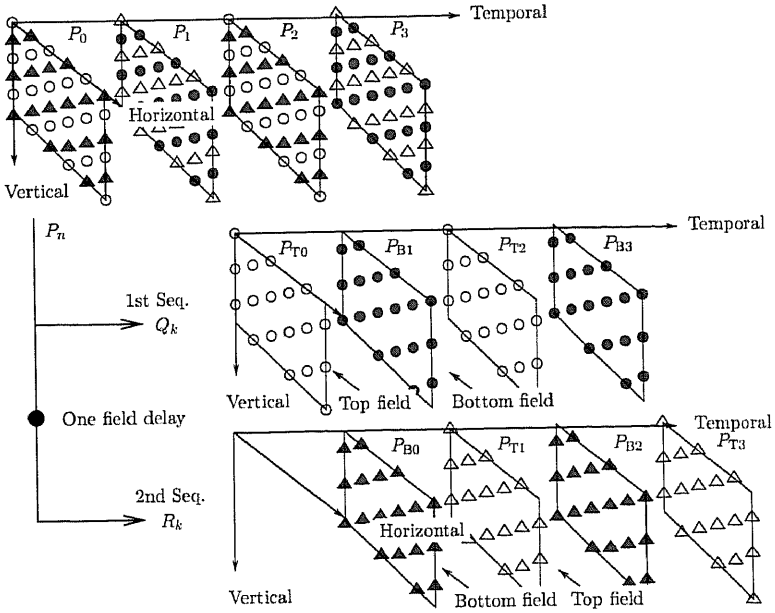
Figure 4.4(a) illustrates an example of line-based invertible deinterlacing and (b) shows the inverse process, i.e. reinterlacing. The invertible deinterlacer constructs a frame picture from two successive top and bottom field pictures as shown in Fig. 4.4 (a), where white, black and gray circles indicate pixels on the top, bottom, and deinterlaced bottom fields, respectively. In this example, the top fields keep their values through the process and a bottom line of a deinterlaced frame picture are obtained by the weighted sum of the three nearest lines with the values alongside the arrows. This process suppresses the comb-tooth artifacts, while maintaining the PR property. The invertible deinterlacing is not restricted to line-based field sequences and we can successfully apply this technique to point-based field sequences [13]. Figure 4.5 (a) illustrates an example of point-based invertible deinterlacing and (b) shows the inverse process. Similarly to the line-based case, the invertible deinterlacer constructs a frame picture from two successive top and bottom field pictures and the pixels on top fields keep their values through the process. Each pixel on bottom fields of a deinterlaced frame picture, in this example, is obtained by the weighted sum of the five nearest pixels with the values alongside the arrows. In this case, dot-pattern artifacts are suppressed instead of comb-tooth artifacts.

4.3.4 Deinterlacer Banks

Figure 4.6 (a) shows a single stage structure of a deinterlacer bank. The deinterlacer bank is composed of a field splitter, two invertible deinterlacers and a frame predictor. The frame predictor is exactly the same operation as the prediction step of lifting wavelet. Existing MCTF schemes consist of a frame split, prediction and update step with motion compensation for the temporal decomposition. In our deinterlacer banks,

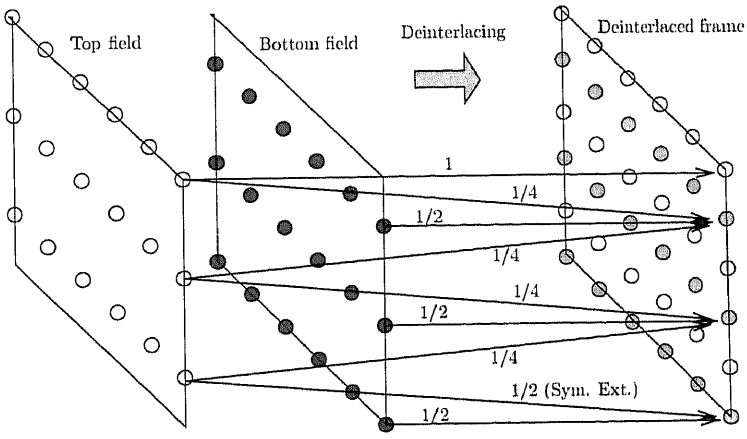


(a) Line-based field split

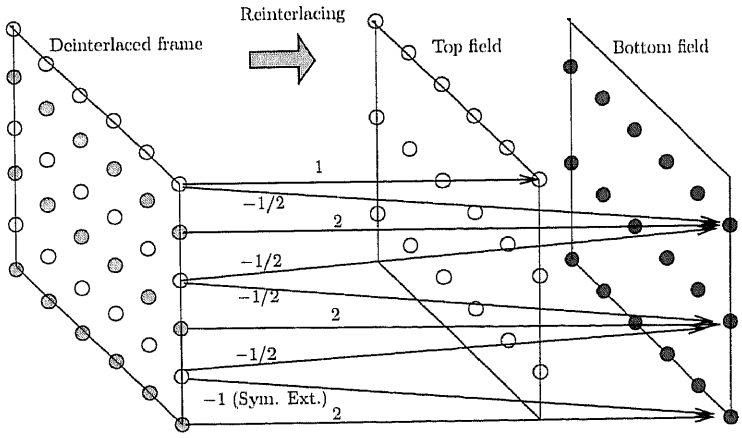


(b) Point-based field split

Figure 4.3: Spatio-temporal split process

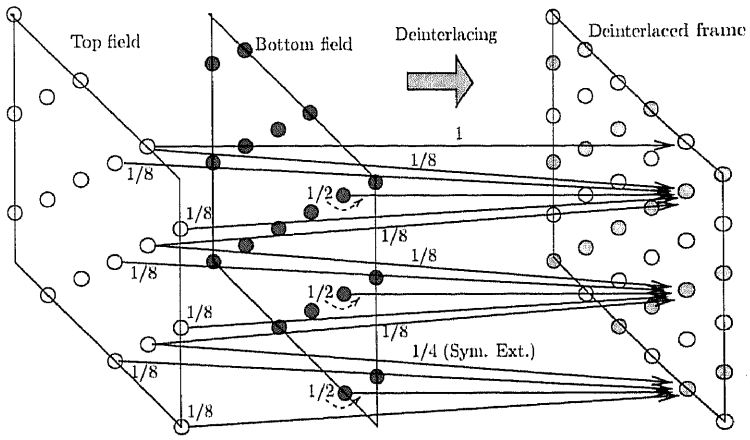


(a) Line-based deinterlacer

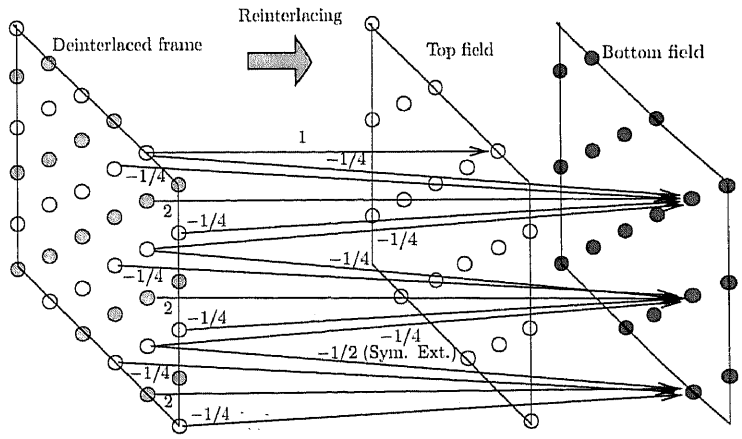


(b) Line-based reinterlacer

Figure 4.4: An example of line-based invertible deinterlacer

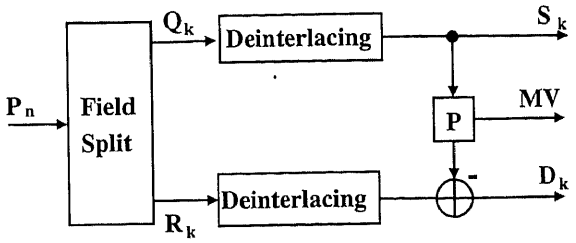


(a) Point-based deinterlacer

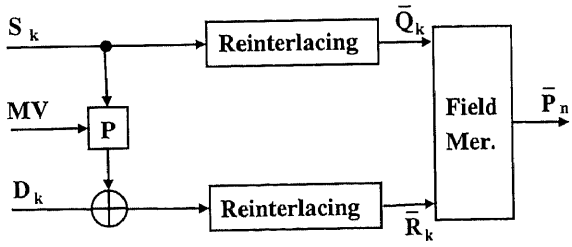


(b) Point-based reinterlacer

Figure 4.5: An example of point-based invertible deinterlacer



(a) Deinterlacer bank



(b) Reinterlacer bank

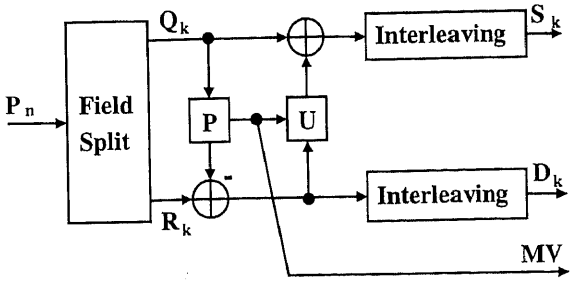
Figure 4.6: Structures of deinterlacer bank and its inverse.

the field split process takes the place of the frame split process. Additionally, our structure disables the update step. This is because the reductions of the complexity due to handling and generating motion vectors in the update operation is expected and ghosting artifacts can be avoided. The main reason is that recovering one of field sequences only from the temporal-low subband sequence becomes possible.

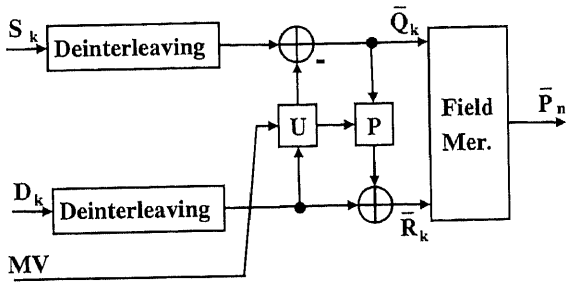
Figure 4.6 (b) shows the inverse transform, which we refer to as reinterlacer bank, where 'Field Mer.' denotes the field merge process, i.e. the inverse process of field splitting. In the field splitting process, input frame sequence P_n is divided into two field sequences Q_k and R_k, where subscripts n and k denote the frame numbers of the full and half frame rate, respectively [44]. The deinterlacer bank finally separates an input frame sequence P_n into two different frame sequences S_k and D_k of a half frame rate, where disagreeable artifacts such as comb-tooth or dot-pattern artifacts are suppressed beforehand by the invertible deinterlacers [12, 13, 18, 45]. Frame sequence D_k is given as prediction errors to exploit the temporal redundancy with a motion compensated frame predictor. Notice that the field sequence Q_k is obtained through the reinterlacer as intermediate data in the reinterlacer bank. The technique provides a new option of spatio-temporal scalability with video format control, which is referred to as *field scalability* [44].

4.4 Non-Separable MCSTF

Non-separable motion compensated spatio-temporal filtering (MCSTF) is discussed as an alternative technique. As mentioned above, the frame rate of given video sequences is controlled with MCTF. When decomposed a video sequence with spatial domain MCTF (t+2D), the frame rate and resolution are controlled individually. In addition,



(a) Analysis bank



(b) Synthesis bank

Figure 4.7: Lifting representation of spatial-temporal filtering.

only videos in the progressive scanning manner are handled. To provide the control of video formats between progressive and interlaced videos, or other non-rectangular sampled video as spatio-temporal scalability, non-separable MCSTF is suggested as an alternative structure. The proposed technique introduces the spatio-temporal split process instead of temporal split used in MCTF. As well, both of line- and point-based field sequences are assumed here.

Figure 4.7 shows the proposed technique. The MCSTF method consists of a spatio-temporal split process (i.e. a field split), a prediction step, an update step, and field interleavers. The interleaver means the operation that merges two successive different fields into a progressive frame picture. MCTF decomposes a progressive video sequence into two subband sequences of a half frame-rate. Even though the system handles interlaced videos as intermediate data, introducing a multi-stage decomposition is simply achieved because the output sequences have also the progressive format as like the input. In the prediction step, MC is applied for decreasing the energy of high-pass subband. The inverse transform is shown in Fig. 4.7 (b), where 'Field Mer.' denotes field merge process, which is the inverse process of the field split process. Hence the system shown in Fig. 4.7 satisfies the PR property, the output frame sequence P_n is completely consistent with the input frame sequence P_n except the delay in the condition without any process to subband signal S_k and D_k. In this study, the proposed structure without the update step is used.

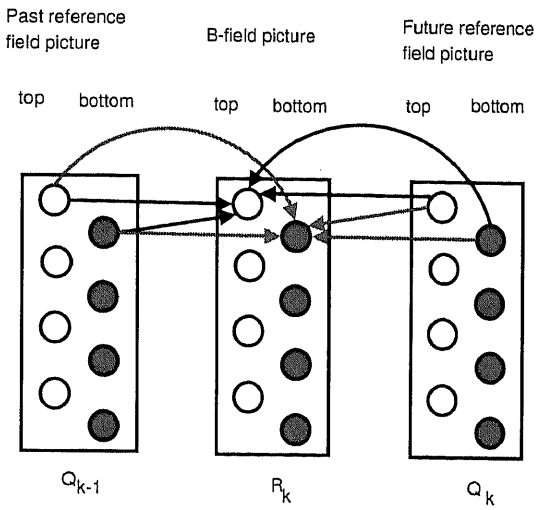


Figure 4.8: Field prediction of field pictures for MCSTF.

4.4.1 Field Motion Compensation

For the proposed structure, the two split field sequences are predicted with MC. And then the low-pass and high-pass subbands are generated. In the prediction step, bidirectional field MC techniques are used. Figure 4.8 shows the procedure of a bidirectional field MC. In this study, the field prediction for field pictures in accordance with MPEG-2 is assumed. In Fig. 4.8, for B-field pictures, the two fields of the two reference pictures are used. For instance, in the case of forward prediction, either top field or bottom field can be selected as a reference field picture to generate the best B-field picture to yield the smallest prediction error (e.g. sum of absolute differences). For backward prediction, similar manner is applied. For the process, four motion vectors to the top and bottom B-fields are obtained.

4.5 Performance Evaluation

To show the significance of the proposed methods, the coding efficiency is evaluated. A video sequence is decomposed to two subbands with MCTF, deinterlacer bank, and non-separable MCSTF, respectively [55]. Afterward, each subband is quantized with scalar quantization and then is dequantized. Finally, the reconstructed video sequence is obtained with each inverse transform so that the PSNRs are evaluated. Simulation parameters are as follows:

- Single stage without update step is used
- Motion compensation with full-pell full-search block matching is applied
  - Macro block sizes for frame and field predictions are 16x16 and 16x32, respectively
  - Bidirectional frame prediction is used for MCTF and deinterlacer bank

  - Bidirectional field prediction is used for MCSTF

- The entropy coding scalar quantization is used for the quantization [28]
- ITE (The institute of Image Information and Television Engineers of Japan) standard SIF video sequence *'Whale Show'* is used as test sequence. (352 × 240, 30fps,8bpp 150th-600th frames)

4.5.1 Global Optimal Rate Allocation

A decision technique of quantization steps obtained with an optimal rate allocation is described. The quantization and dequantization are expressed as follows.

$$q = Q(x) = \begin{cases} \text{sign}(x)\lfloor \frac{|x|}{\Delta} + \xi \rfloor & \frac{|x|}{\Delta} + \xi \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

$$\hat{x}_q = IQ(q) = \begin{cases} 0 & q = 0 \\ \text{sign}(q)(|q| - \xi + \delta)\Delta & q \neq 0 \end{cases} \tag{4.2}$$

where, δ dedicates the quantization step. ξ and δ are parameters. In this simulation, $\xi = 0$ and $\delta = \frac{1}{2}$ are assumed, respectively. The quantization of subband b depends on the correlations between the transformed coefficients in its subband. Therefore, for subband b, the quantization step δ_b can be individually decided. When full band rate R is assumed, the rate δ_b to which MSE minimized in subband b is obtained. Then, the quantization step δ_b is decided with the rate R_b. As a result, the optimal rate allocation is given by the following equation:

$$R_b = R + \frac{1}{2}\log_2 \frac{G_b W_b \varepsilon_b^2 \sigma_b^2}{\prod_{l=0}^{B-1} (G_l W_l \varepsilon_l^2 \sigma_l^2)^{\eta_l}}, \tag{4.3}$$

where η_b is the fraction bits in the b-th subband to the full band, G_b is decided by the gain of the synthesis filters. W_b is also the constant to give the weighted MSE, $W_b = 1$ is assumed in this simulation. In this experiment, IID Gaussian data is assumed, then $\varepsilon_b^2 = \varepsilon^2 \cong 1$. As well, for variances σ_b^2, each variance obtained by each subband for the test sequence is used. The gains of each synthesis filters listed in Table 4.1. For the gains of MCSTF, the same gain of MCTF can be used since MCSTF has the identical filter coefficients with the synthesis bank. The step size δ_b is given by

$$\Delta_b = \sqrt{\frac{\varepsilon_b^2 \sigma_b^2}{c_b} 2^{-R_b}}, \tag{4.4}$$

where C_b is constant. When ECSQ is used, $C_b = \frac{1}{12}$ is obtained.

4.5.2 Experimental Results

Average PSNR

Figure 4.9 shows the average PSNR resulting from the systems through the MCTF, the deinterlacer banks, and the MCSTF, respectively. Through a comparison between the results of MCTF and the deinterlacer banks in Fig. 4.9, the PSNRs of MCTF are

Table 4.1: Gain G_b of the synthesis filters

	G_0(Low-pass band)	G_1(high-pass band)
MCTF	1.50	1.00
MCSTF	1.50	1.00
Deint. bank(line-based)	2.250 (Top), 6.00(Bottom)	1.50(Top), 4.00(Bottom)
Deint. bank(point-based)	1.875 (Top), 8.00(Bottom)	1.25(Top), 4.00(Bottom)

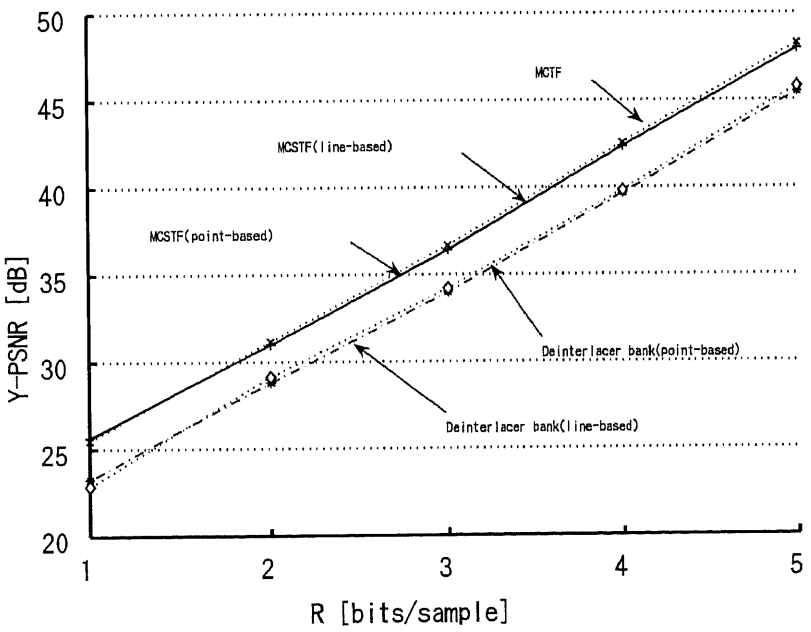


Figure 4.9: Average PSNR with entropy coding scalar quantization with Sequence '*Whale Show*'.

about 2.5 dB higher than that of the deinterlacer banks. Similarly, the results of MCTF shows only 0.4 dB higher performance compared with that of MCSTF. It is verified that such a structure composed field prediction techniques with MC have almost the same performance. For deinterlacer banks, deinterlaced frame pictures are generated after the split process. Then, frame prediction is applied for these frame pictures. As a result, the prediction error increases in the high-pass subband so that the coding performance would be decreased by the increase of the quantization error.

PSNR Fluctuation

Figure 4.10 shows the PSNRs of each frame at average bit-rate $R = 3$ [bpp] with MCTF and MCSTF, respectively. The result from MCTF without the update has the PSNR fluctuations between the even number and the odd number frame as shown in Fig. 4.10. In contrast, both of the MCSTF results suppress the fluctuations even if it doesn't use

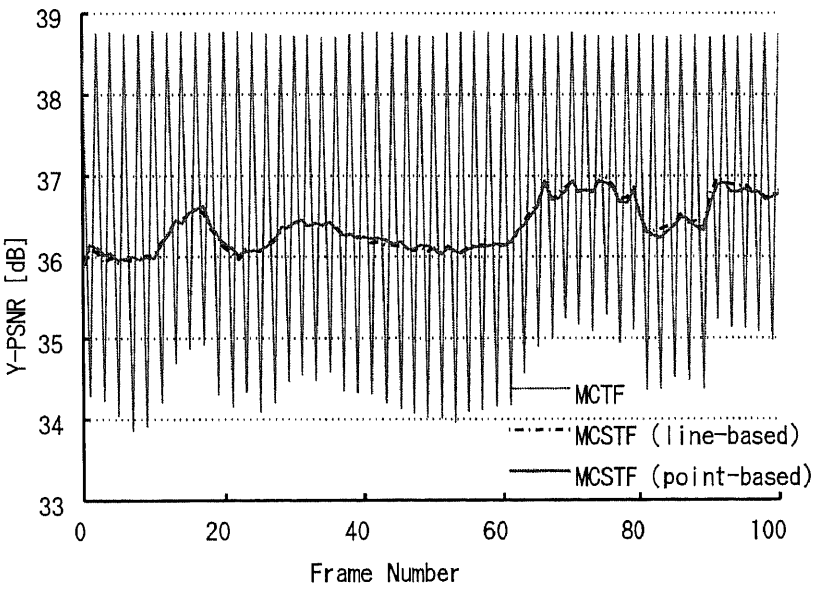
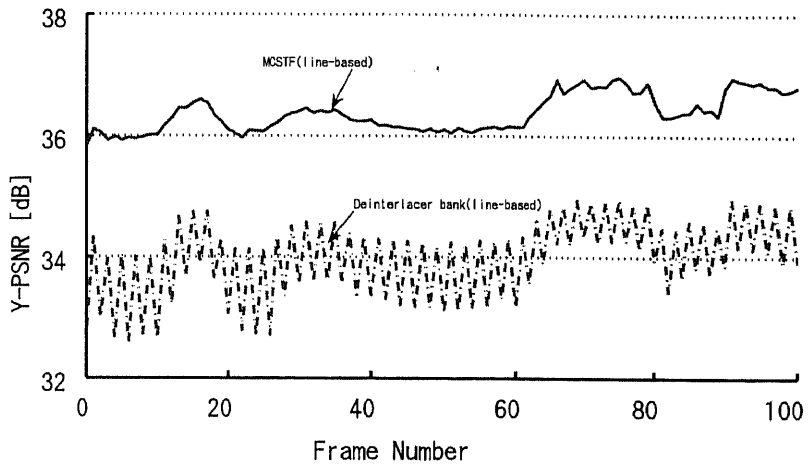


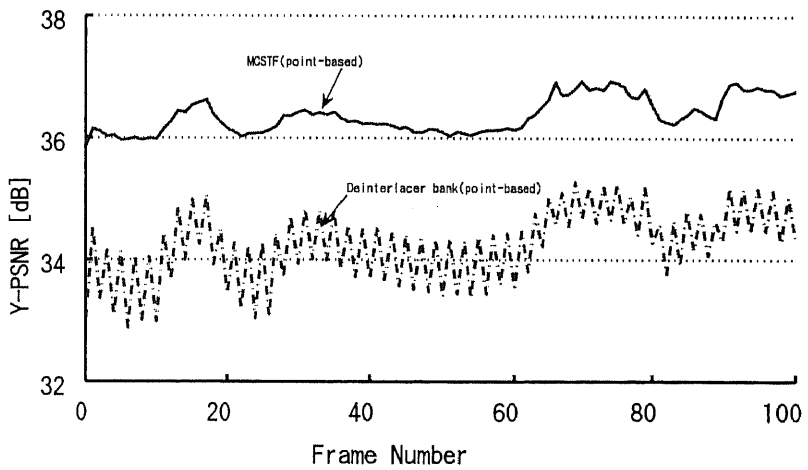
Figure 4.10: PSNR of MCTF, MCSTF(line-based), and MCSTF(point-based) in the case of R=3[bpp], where Sequence '*Whale Show*' is used.

the update step when compared with MCTF. In MCSTF, the sampling values of frame pictures are divided into the two field pictures through the spatio-temporal split process in Fig. 4.3. The interleaved frame pictures resulting from the split process in Fig. 4.7 has the sampling values of the adjoined two original frame pictures half, respectively. Therefore, in all frames reconstructed with the scalar quantization, either sampling values of the even number or the odd number frame of the original video sequence must be maintained so that image quality is kept with stability. For the simulation, with a single stage structure, it is verified that the differences of PSNR between the even and odd frames is reduced through the MCSTF structure without the update step when compared with the result from MCTF.

As well, Figure 4.11 shows the comparison with the deinterlacer banks and MCSTF. From the results, the deinterlacer banks can suppress the PSNR fluctuation. These results show that the simple structure expect for the update step enables the suppression of the fluctuations by introducing the spatio-temporal spilt process. As shown in Fig. 4.11, MCSTF has higher performance in terms of the fluctuations than the deinterlacer banks. It is verified that MCSTF composed of field prediction techniques with MC performs efficient coding and suppresses the differences between the frames as shown in Fig. 4.11 while reducing redundancy in video sequences for temporal direction. It is verified that better energy compaction can be achieved by introducing field prediction with MC, while keeping the suppression of the PSNR fluctuations.



(a) Line-based techniques



(b) Point-based techniques

Figure 4.11: Comparison between deinterlacer bank and MCSTF in the case of R=3[bpp], where Sequence '*Whale Show*' is used.

4.6 Concluding Remarks

In this chapter, novel systems through spatio-temporal analysis for scalable video coding were presented. To provide the control between progressive and non-rectangular sampling video sequences as a new tool, two types of techniques handling field sequences as intermediate data were proposed: deinterlacer banks and motion compensated spatio-temporal filtering systems. The VT-quincunx and FCO format were dealt with as interlaced video formats. These proposed techniques bring a new spatio-temporal scalability that provides progressive and interlaced scanning video using the field split process according with these VT-quincunx and FCO format, respectively.

Firstly, for one of the proposal techniques, deinterlacer banks were presented. Unlike the conventional filter banks, the systems with deinterlacer banks are constructed in a way unique to multidimensional systems using invertible deinterlacers.

For an alternative approach, a system with non-separable motion compensated spatio-temporal filtering was suggested. The structures introducing field prediction with MC techniques through the spatio-temporal split process were discussed. In the simulation, the coding performance was evaluated with single stage structures without the update step for MCTF, deinterlacer banks and MCSTF, respectively. The suppression of the PSNR fluctuations can be achieved with the proposed methods without the update step. In particular, it was verified that MCSTF has higher suppression effect than the others. The interleaving process with MCSTF, however, has a problem that the simple interleaving generates comb-tooth or dot-pattern artifacts in low-pass subband through MCSTF. The deinterlacing technique can be applied to suppress these artifacts for the low-pass subband. Therefore, it is necessary to investigate field prediction structures with the invertible deinterlacer.

Chapter 5

Conclusion

5.1 Summary

This thesis dealt with non-separable multirate systems with invertible sampling lattice conversion for scalable video coding. The contribution of the thesis is to have given the coding system for intra-based coding application with the invertible deinterlacing techniques, and to have provided the format control between progressive scan and interlaced scan formats using the non-separable multi-rate systems with spatio-temporal decomposition as new scalable functionalities. The summarizations are as follows:

Chapter 2 has presented lossy compression of a Motion-JPEG2000 (MJP2) codec system with invertible deinterlacing. The invertible video format conversion has been developed a new class of deinterlacing. The deinterlacing technique has the sampling density preservation and the perfect reconstruction property. The invertible deinterlacer can suppress comb-tooth artifacts which are caused by field interleaving for interlaced scanning videos and affect the quality of scalable intraframe-based codecs such as MJP2. The main point of this study is subjected the implementation issues. The application scenario with this technique as pre-filter was suggested. Since the invertible deinterlacer can be implemented as a lifting structure, it is possible to combine this lifting-based deinterlacer with DWT in MJP2. In addition, to improve the quality of recovery with inverse process (i.e. reinterlacing), an energy gain compensation technique was discussed using the MJP2 system integrated with the invertible deinterlacer. It was verified that the quality recovery can be improved through the compensation for the energy gain factors of subbands by taking the influence of reinterlacing into account. The simulation results showed that the SNR scalability of the frame-based MJP2 coding system for interlaced scanning videos was maintained and the proposed gain compensation achieved better quality recovery than the default gain, which still keeps the filtering effect for suppressing the comb-tooth artifacts within the standard MJP2 bitstream format. It was found that the comb-tooth suppression capability increases while deinterlacing filter becomes close to a vertical filter. In contrast, the quality recovery was gained much more while a filter became close to a temporal filter,

Chapter 3 has discussed a lossless implementation of Motion-JPEG2000 integrated with invertible deinterlacing. When the structure combined the invertible deinterlacer with the integer transform in MJP2 is directly applied, fractions in high frequency subbands are generated due to the scaling process of the deinterlacer. In the standard MJP2 encoder, the resulting fractions are not taken into account in terms of determining the

number of the bit-depth of the transform coefficients since the bit-depth is decided according as subband gain factors defined in MJP2. In order to avoid discarding fractions generated due to the scaling, we can give the compensated gains instead of the original gains because the scaling factor can be obtained as the power of two. As an exact implementation, it was shown that the lossless compression can be realized by recalculating the subband gain factors and modifying the marker segments. With several experimental results, it was verified that the coding efficiency was almost the same as that obtained by the normal field interleaving, while maintaining the suppression capability of comb-tooth artifacts with the standard decoding.

In Chapter 4, Novel systems providing a spatio-temporal functionality for scalable video coding were presented. To provide the control between rectangular and non-rectangular sampling video sequences as a new tool, two types of techniques handling field sequences as intermediate data were proposed: deinterlacer banks and motion compensated spatio-temporal filtering systems. The VT-quincunx and FCO format were used as interlaced video formats. These proposed techniques bring a new spatio-temporal scalability that provides the progressive and interlaced scanning video using the field split process according with these VT-quincunx and FCO format, respectively. The proposed systems separate a progressive video into two different progressive videos of a half frame rate and enable us to handle interlaced videos are given as intermediate data during analysis and synthesis process. It was verified that the technique can provide the video format control between progressive and interlaced scan formats as a new spatio-temporal scalability. In this thesis, deinterlacer banks were presented as the first approach. Unlike the conventional filter banks, the systems with deinterlacer banks are constructed in a way unique to multidimensional systems using invertible deinterlacers. For an alternative approach, a system with non-separable motion compensated spatio-temporal filtering was suggested. The structures introducing field prediction with motion compensation techniques were discussed.

In simulation results, the coding performance was evaluated with single layer structures without the update step for MCTF, deinterlacer banks and MCSTF, respectively. The suppression of the PSNR fluctuations can be achieved with the proposed methods without the update step. In particular, it was verified that MCSTF has higher suppression effect than the others.

5.2 Open Problems

There remain some important problems for the proposed MCSTF as non-separable multirate systems. The interleaving process with MCSTF causes comb-tooth or dot-pattern artifacts in the low-pass subband. The deinterlacing technique can be applied to suppress these artifacts. Therefore, it needs to consider field prediction structures with the invertible deinterlacer. As well, it is of interest to investigate structures combined 2-dimensional discrete wavelet transform as a spatial transform. On the other hand, in order to achieve local adaptability, it is necessary to introduce variable-coefficient invertible deinterlacers [15] into MCSTF, and to examine spatio-temporal multi-resolution analysis using hierarchical structures composed of MCSTF and wavelet transform as a spatial decomposition and its performance evaluation. Furthermore, it is also important to estimate the performance of these spatio-temporal multi-resolution systems using a rate-distortion control and entropy coding.

Bibliography

[1] A. Murat Tekalp, *Digital Video Processing*, Prentice Hall, Inc., 1995.

[2] G. de Haan and E. B. Bellers, "Deinterlacing-an overview," in *Proc. IEEE*, Sept. 1998, vol. 86, pp. 1837–1857.

[3] E. B. Bellers and G. de Haan, *De-interlacing: A Key Technology for Scan Rate Conversion (Advances in Image Communication, 9)*, North-Holland, 2000.

[4] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.*, Prentice Hall, Englewood Cliffs, 1993.

[5] S. Muramatsu and H. Kiya, "Parallel Processing Techniques for Multidimensional Sampling Lattice Alteration Based on Overlap-add and Overlap-save Methods," *IEICE Trans. Fundamentals*, vol. E78-A, no. 8, pp. 934–943, Aug. 1995.

[6] L. Vandendorpe and L. Cuvelier, "Statistical Properties of Coded Interlaced and Progressive Image Sequences," *IEEE Trans. Image Processing*, Vol.8, no.6, pp.749–761, June 1999.

[7] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications.* Prentice-Hall, 2002.

[8] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards Algorithms and Architectures Second Edition*, Kluwer Academic Publishers, 1997.

[9] T. Kuge and H. Hoshino, "A Study of HDTV Moving Picture Coding Using the JPEG2000," *Proc. IEICE Spring Conv.*, D11-18, p.18, Mar. 2001.

[10] T. Kuge, "Wavelet picture coding and its several problems of the application to the interlace HDTV and the ultra-high definition images," in *IEEE Proc. of ICIP*, WA-P2.1, 2002.

[11] A. Basso, G. Cash and M. Civanlar, "Transmission of MPEG-2 Streams over Non-Guaranteed Quality of Service Networks," *Picture Coding Symposium (PCS 97)*, Berlin, Germany, September, 1997.

[12] S. Muramatsu, T. Ishida, and H. Kikuchi, "A design method of invertible deinterlacer with sampling density preservation," in *IEEE Proc. of ICASSP*, vol. 4, pp. 3277–3280, May 2002.

[13] S. Muramatsu T. Ishida and H. Kikuchi, "Invertible deinterlacer with sampling-density preservation: Theory and design," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2343–2356, Sept, 2003

[14] H. Kikuchi, S. Muramatsu, T. Ishida, and T. Kuge, "Reversible conversion between interlaced and progressive scan formats and its efficient implementation," *Proc. of EUSIPCO*, no. 448, pp. 275–278, 2002.

[15] T. Ishida, S. Muramatsu H. Kikuchi and T. Kuge, "Invertible deinterlacing with variable coefficients and its lifting implementation," in *IEEE Proc. of ICASSP*, IMSP-L5.3(III-105), April 2003. (also appears in *IEEE Proc. of ICME, ICASSP-8.7(III-177), July 2003*)

[16] T. Ishida, T. Soyama S. Muramatsu H. Kikuchi and T. Kuge, "A lifting implementation of variable-coefficient invertible deinterlacer with embedded motion detector," *IEICE Trans. Fundamentals*, vol. E86-A, pp. 1942–1948, Aug. 2003.

[17] T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, and H. Kikuchi, "DWT Gain Compensation of Motion-JPEG2000 for an Invertible Deinterlacer," in *IEEE Proc. ICIP*, No.TA-P1.12, 2003.

[18] T. Ishida, S. Muramatsu, and H. Kikuchi, "Motion-JPEG2000 Codec compensated for Interlaced Scanning Videos," *IEEE Trans. Image Processing*, vol.14, no.12, pp.2179-2191, Dec. 2005.

[19] T. Soyama, T. Ishida S. Muramatsu H. Kikuchi and T. Kuge, "Lifting architecture of invertible deinterlacing," *IEICE Trans. Fundamentals*, vol. E86-A, no. 4, pp. 779–786, Apr. 2003.

[20] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.

[21] Y. Harada, S. Muramatsu, and H. Kiya, "Multidimensional multirate filter and filter bank without checkerboard effect," *Proc. of EUSIPCO*, pp. 1881–1884, Sept. 1998.

[22] H. Kiya, K. Nishikawa, and M. Iwahashi, "A development of symmetric extension method for subband image coding," *IEEE Trans. Image Processing*, vol. 3, no. 1, pp. 78–81, Jan. 1994.

[23] Cannon, EPFL and Ericsson "http://jj2000.epfl.ch,"

[24] T. Fukuhara, K. Katoh, S. Kimura, K. Hosakam, and A. Leung, "Motion-JPEG2000 Standardization and Target Market," in *IEEE Proc. ICIP*, TA02.08, pp. II57–II60, 2000.

[25] ITU-T T.800 Recommendation, "Information technology - JPEG2000 image coding system : Core coding system" Oct., 2002.

[26] ISO/IEC FDIS 15444-5 Final Draft International Standard, "Information Technology - JPEG2000 images coding system : Reference software", 2001.

[27] ISO/IEC JTC1/SC 29/WG1 N2117, "Motion JPEG2000 Final Committee Draft 1.0," March 2001.

[28] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, 2002.

[29] D. Marpe, V. George, H. L. Cycon, and K. U. Barthel, "Performance Evaluation of Motion-JPEG2000 in Comparison With H.264/AVC Operated in Pure Intra Coding Mode", *Proc.SPIE on Wavelet Applications in Industrial Processing*, Photonics East, Rhode Island, Oct. 2003.

[30] L. Vandendorpe, L. Cuvelier, B. Maison, and P. Delogne, "Coding of deinterlaced image sequences," *IEEE Proc. ICIP*, pp.943–946, Nov. 1994.

[31] L. Vandendorpe, L. Cuvelier, and B. Maison, "Human Visual Weighted Quantization for Transform/Subband Image Coding Revisited for Interlaced Pictures," *IEEE Trans. Image Processing*, vol. 7, no. 2, pp. 222–225, Feb. 1998

[32] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression", *IEEE Trans. on Image Processing*, vol. 12, no. 12, pp. 1530–1542, 2003.

[33] J.-R.Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[34] S.-J. Choi, J. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.

[35] B. Pesquet-Popescu, and V. Bottreau,"Three-dimensional lifting schemes for motion compensated video compression," *ICASSP*, vol. 3, pp. 1793–1796, Salt Lake City, 2001.

[36] J. Xu, Z. Xiong, S. Li, Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimal truncation (3D-ESCOT)," Applied and Computational Harmonic Analysis, 10 (2001) 290-315.

[37] L. Luo, J. Li, S. Li, Z. Zhuang, and Y.-Q. Zhang, "Motion compensated lifting wavelet and its application in video coding," *ICME*, pp. 365–368, Tokyo, 2001.

[38] N. Mehrserest and D. Taubman, "Adaptively weighted update steps in motion compensated lifting based scalable video compression, " *IEEE Proc. ICIP*, vol. 2, pp. 771–774, Sept. 2003.

[39] R. Xiong, F. Wu, S. Li, Z. Xiong and Y.-Q. Zhang, "Exploiting temporal correlation with adaptive block-size motion alignment for 3D wavelet coding," *SPIE*, vol. 5308, pp. 144–155, 2004.

[40] C. Tiller, B.Pesquet-Popescu, and M.van der Schaar, "Weighted average spatio-temporal update operator for subband video coding," *IEEE Proc. ICIP*, pp.1305–1308, 2004.

[41] F. Verdicchio, Y. Andreopoulos, T. Clerckx, J. Barbarien, A. Munteanu, J. Cornelis, and P.Schelkens, "Scalable video coding based on motion-compensated temporal filtering: Complexity and functionality analysis," *IEEE Proc. ICIP*, pp.2845–2848, 2004.

[42] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis and P. Schelkens, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication* (special issue on "Subband/Wavelet Interframe Video Coding"), vol. 19, no. 7, pp. 653–673, Aug. 2004.

[43] Y. Andreopoulos, A. Munteanu, G. Van der Auwera, J. Cornelis and P. Schelkens, "Complete-to-overcomplete discrete wavelet transforms: theory and applications," *IEEE Trans. on Signal Processing*, vol. 53, no. 4, pp. 1398–1412, April 2005.

[44] S. Muramatsu, T. Ishida, and H. Kikuchi,"Perfect reconstruction deinterlacer banks for field scalable video compression," *IEEE ICIP2004*, pp.2279-2282, Oct. 2004.

[45] D. Kitagawa, T. Ishida, S. Muramatsu and H. Kikuchi: "Maximally-decimated perfect reconstruction deinterlacer banks Based on FCO-Sampling," *ITC-CSCC*, pp.7B1L-3-1 – 7B1L-3-4, July 2004.

[46] R. Calderbank, I. Daubechies, W. Sweldens, and B.-L.Yeo, "Wavelet transforms tha map integers to integers," *Appl.Comput.Harmon.Anal.*, vol. 5, pp. 332–369, July 1998.

[47] J.-R.Ohm, "Advances in scalable video coding,"*Processing of the IEEE*, vol. 93, no. 1, pp.42–56, Jan 2005.

[48] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*, The IEE, 2003.

[49] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 11, pp.301–317, Mar. 2001.

[50] R. K. Rajebdran, M.van der Schaar, and S.-F. Chang, "FGS+: Optimizing the joint SNR-temporal video quality in MPEG-4 fine grained scalable coding," *IEEE Proc. ISCAS*, vol. 1, pp.445-448, 2002.

[51] R. Schafer, H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "Mctf and Scalability Extension of H.264/AVC and Its Application to Video Transmission, Storage, and Surveillance," in *Proc. SPIE, Visual Commun. Image Process.*, pp.343–354, Beijing, China, July 2005.

[52] ISO/IEC JTC1/SC29/WG11, "Joint Scalable Video Model jsvm-4," Document, Oct. 2005.

[53] L. Song, J. Xu, H.Xiong, and F. Wu, "Content adaptive update steps for lifting-based motion compensated temporal filtering," *Proc. Picture Coding Symp.*, San Francisco, CA, USA, Dec. 2004.

[54] J. Kovacevic and M. Vetterli, "FCO Sampling of Digital Video Using Perfect Reconstruction Filter Bank," *IEEE Trans. on Image Processing*, vol. 2, no. 1, pp.118–122, Jan. 1993.

[55] OpenFiesta; "URL: http://viplab.eng.niigata-u.ac.jp/openfiesta/".

Research Works

Academic Papers

Full Papers

1. T. Ishida, S. Muramatsu and H. Kikuchi: "**Motion-JPEG2000 Codec Compensated for Interlaced Scanning Videos**," *IEEE Trans. on Image Processing, Vol. 14, No. 12, pp. 2179–2191, Dec. 2005*

2. T. Ishida, S. Muramatsu and H. Kikuchi: "**Lossless Compression of Motion JPEG2000 Integrated with Invertible Deinterlacing**," *ITE, Vol. 59, No. 7, pp. 1011–1019, July 2005*

3. S. Muramatsu, T. Ishida, and H. Kikuchi: "**Invertible Deinterlacing with Sampling Density Preservation:Theory and Design**," *IEEE Trans. on Signal Processing, Vol. 51, No. 9, pp. 2343–2356, Sep. 2003*

4. T. Ishida, T. Soyama, S. Muramatsu, H. Kikuchi, and T. Kuge: "**A lifting Implementation of Variable-Coefficient Invertible Deinterlacer with Embedded Motion Detector**," *IEICE Trans. on Fundamentals, Vol. E86-A, No. 8, pp. 1942–1948, Aug. 2003*

5. T. Soyama, T. Ishida, S. Muramatsu, H. Kikuchi, and T.Kuge: "**Lifting Architecture of Invertible Deinterlacing**," *IEICE Trans. on Fundamentals, Vol. E86-A, No. 4, pp. 779-786, Apr. 2003*

Letter

1. T. Ishida,S. Muramatsu, H. Kikuchi, and T. Kuge: "**Invertible Deinterlacing with Variable Coefficients**," *IEICE Trans. on Fundamentals, Vol. J87-1, No. 2, pp. 336–342, Feb. 2004 (in Japanese)*

International Conferences

1. T. Ishida, S. Muramatsu, D. Kitagawa, J. Uchita, M. Hiki, H. Kikuchi: "**Performance Evaluation of Spatio-Temporal Multi-Resolution Analysis with Deinterlacer Banks**," *Visual Communications and Image Processing, Proc. of SPIE, vol. 5960, pp. 177–188, Beijing, July 2005*

2. T. Ishida, S. Muramatsu, H. Kikuchi: "**Lossless Implementation of Motion-JPEG2000 Integrated with Invertible Deinterlacing,**" *IEEE International Symposium on Circuits and systems(ISCAS), C4P-Y.6, pp.6328–6331, Kobe, May 2005*

3. T. Ishida, S. Muramatsu, J. Uchita, D. Kitagawa, H. Kikuchi: "**Spatio-Temporal Multi-Resolution Analysis through Hierarchical Deinterlacer Banks,**" *International Workshop on Advanced Image Technology(IWAIT), pp. 139–144, Cheju, Korea, Jan. 2005*

4. T. Ishida, J. Uchita, S. Muramatsu, H. Kikuchi, T. Kuge: "**Parameter Transmission via ROI in JPEG2000 for Variable-Coefficient Invertible Deinterlacing,**" *Proc. EUSIPCO, No. 1251, pp. 653-656, Vienna, Austria, Sep. 2004*

5. D. Kitagawa, T. Ishida, S. Muramatsu and H. Kikuchi: "**Maximally-decimated perfect reconstruction deinterlacer banks Based on FCO-Sampling,**" *pp. 7B1L-3-1 – 7B1L-3-4, July 2004*

6. J. Uchita, S. Muramatsu, T. Ishida, H. Kikuchi, T. Kuge: "**Parameter Embedding Method of Variable-Coefficient Invertible Deinterlacing into Motion-JPEG2000 through ROI,**" *IEEE Mid-West Symp. on Circuits and Systems 2004, pp. II425–II428, Hiroshima, July 2004*

7. S. Muramatsu, T. Ishida, and H. Kikuchi: "**Perfect reconstruction deinterlacer banks for field scalable video compression,**" *IEEE ICIP2004, pp. 2279-2282, TP-P7.8, Singapore, Oct. 2004*

8. T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, and H. Kikuchi: "**DWT Gain Compensation of Motion-JPEG2000 for an Invertible Deinterlacer,**" *IEEE Proc. ICIP, No. TA-P1.12, Barcelona, Spain, Sep. 2003*

9. J. Uchita, T. Ishida, S. Muramatsu, H. Kikuchi, T. Kuge: "**Parameter Decimation for Invertible deinterlacing with Variable Coefficients,**" *Proc. of International Technical Conference on Circuits/Systems, Computers and Communications, Phenix Park, Kan-Won, Korea, pp. 925–928, July 2003*

10. T. Ishida, S. Muramatsu H. Kikuchi and T. Kuge: "**Invertible deinterlacing with variable coefficients and its lifting implementation,**" *IEEE Proc. of ICASSP, IMSP-L5.3(III-105), April 2003. (also appears in Proc. of ICME, ICASSP-8.7(III-177), July 2003)*

11. H. Kikuchi, S. Muramatsu, T. Ishida, and T. Kuge: "**Reversible conversion between interlaced and progressive scan formats and its efficient implementation,**" *Proc. of EUSIPCO, no. 448, pp. 275–278, Toulouse, France, Sep. 2002*

12. T. Ishida, T. Soyama, S. Muramatsu, H. Kikuchi, T. Kuge: "**Lifting Implementation of Reversible Deinterlacer,**" *Proc. of International Technical Conference on Circuits/Systems, Computers and Communications, vol.1, pp. 90–93, Phuket, Thailand, July 2002*

Domestic Conferences

1. T. Ishida, M. Hiki, S. Muramatsu, H. Kikuchi: "**Non-Separable Motion Compensated Spatio-Temporal Filtering for Scalable Video Coding,**" *IEICE Proc. 20 th Digital Signal Processing Symp., D5-2, Nov. 2005 (in Japanese)*

2. T. Ishida, S. Muramatsu, M. Hiki, H. Kikuchi: "**Hierarchical Variable-Coefficient Deinterlacer Banks for Scalable Video Coding,**" *ITE Information·Media Rept., vol. 29, no. 34, pp. 9–12, June 2005 (in Japanese)*

3. T. Ishida, S. Muramatsu, J. Uchita, D. Kitagawa, H. Kikuchi: "**Hierarchical Deinterlacer Banks for Spatio-Temporal Scalable Video Coding,**" *19th Picture Coding Symposium of Japan, pp. 35–36, Nov. 2004 (in Japanese)*

4. T. Ishida, S. Muramatsu, D. Kitagawa, S. Kobayashi, H. Kikuchi: "**Perfect Reconstruction Deinterlacer Banks for Field Scalable Video Coding,**" *IEICE 17th Circuits and Systems Karuizawa Workshop, pp. 469–474, Apr. 2004 (in Japanese)*

5. T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, H. Kikuchi: "**Lossless implementation of Motion-JPEG2000 integrated with invertible deinterlacer,**" *IEICE 18th Picture Coding Symposium of Japan, pp. 11–12, Dec. 2003 (in Japanese)*

6. T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, H. Kikuchi: "**Evaluation of Filters in Invertible Deitnerlacer for Motion-JPEG2000,**" *ITE Rept. vol. 27, no. 34, pp. 13–16, June 2003 (in Japanese)*

7. T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, H. Kikuchi: "**Frequency Weight Adjustment of Motion-JPEG2000 for Invertible deinterlacer,**" *IEICE 16th Circuits and Systems Karuizawa Workshop, pp. 387–391, Apr. 2003*

8. S. Muramatsu, T. Ishida, T. Soyama, H. Kikuchi, T. Kuge: "**Lifting Implementation of Variable-Coefficient Invertible Deinterlacer,**" *IEICE 17th Picture Coding Symposium of Japan, pp. 97–98, Dec. 2002 (in Japanese)*

9. S. Muramatsu, T. Ishida, T. Soyama, H. Kikuchi, T. Kuge: "**Efficient Implementation of Variable-Coefficient Invertible Deinterlacing,**" *IEICE Proc. 17th Digital Signal Processing Symp., Dec. 2002 (in Japanese)*

10. T. Soyama, T. Ishida, S. Muramatsu. H. Kikuchi, T. Kuge: "**Lifting Implementation of Perfect Reconstruction Deinterlacing,**" *IEICE 15th Circuits and Systems Karuizawa Workshop, pp. 89–94, Apr. 2002 (in Japanese)*

11. S. Muramatsu, T. Ishida, H. Kikuchi, T. Kuge: "**Variable Coefficients Invertible Deinterlacing,**" *IEICE Rept., No. CAS2002-33, pp. 43–48, June 2002 (in Japanese)*

12. T. Ishida, S. Muramatsu, J. Zhou, S. Sasaki, H. Kikuchi: "**Frame/Field Intra-Motion Picture Coding with Perfect Reconstruction Deinterlacing ,**" *IEICE Rept., No. CAS2001-80, pp. 25–32, Dec. 2001 (in Japanese)*