PERFECT RECONSTRUCTION DEINTERLACER BANKS FOR FIELD SCALABLE VIDEO COMPRESSION

Shogo Muramatsu, Takuma Ishida, and Hisakazu Kikuchi

Dept. of Electrical and Electronic Eng., Niigata University
8050, 2-no-cho Ikarashi, Niigata, 950-2181, Japan

ABSTRACT

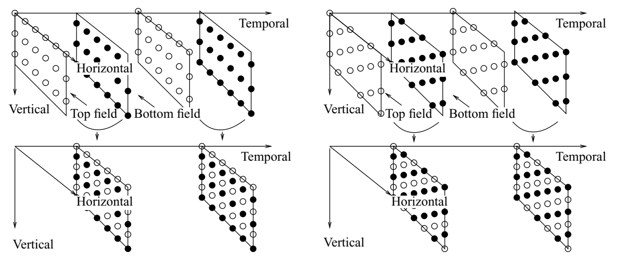
Perfect reconstruction (PR) deinterlacer banks are proposed as new tools for spatio-temporal scalable video codec. The proposed systems separate a progressive video into two different progressive videos of a half frame rate and are novel from the viewpoint of filter-banks in that interlaced videos are 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 paper suggests two kinds of interlaced video formats: the vertical-temporal (VT) Quincunx and face-centered-orthorhombic (FCO) format. As a primal application, a spatio-temporal scalable video codec, or *field scalable video codec (FSVC)*, system is experimented. This technique offers a functionality by which decoding a base layer by itself provides both of half-rate progressive and interlaced videos, and adding an enhancement layer improves the spatio-temporal resolution. Some experimental results of the filter banks combined with JPEG2000 show their potential for practical video codec applications.

1. INTRODUCTION

The ideal goal of scalable video compression is to encode a video once and then partially decode the bit-stream by truncating 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 Internet video streaming requires 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 videos by gradually 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 [1]. 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 [2], 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 [3]. Independently from standards, there are some trials to make the scalability much finer. In the articles [4] and [5], a three-dimensional (3-D) discrete wavelet transform (DWT) is introduced, in which a spatio-temporal transform precedes embedded quantization and bit-plane coding as defined in the current state-of-the-art scalable image codec JPEG2000 [6]. In the 3-D DWT embedded video coding, filter banks play an important role for the spatial-temporal scalability.

Part of this work was developed during Muramatsu's stay in Florence, Italy as a visiting scientist. We would like to thank Prof. Alberto Del Bimbo, Dept. of System & Information Eng., University of Florence, Italy, for his kind support and encouragement.

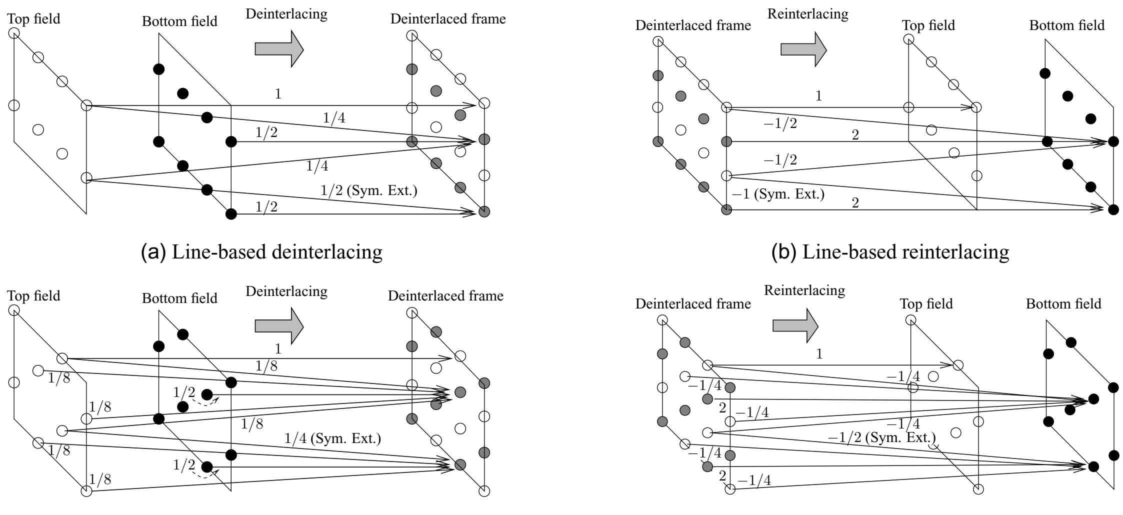


(a) VT Quincunx (Line base) (b) FCO (Point base)

Fig. 1: Video formats before (on top) and after (on bottom) deinterlacing, where white and black circles are pixels on top and bottom fields, respectively.

Until now, we have two choices of the spatio-temporal resolution control with respect to video formats. One handles a video only in the progressive scanning manner [4, 5], and the other divides a video into an interlaced, or non-rectangular sampled, field sequence and its residual [1, 7]. 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 this work, we introduce a new option to the spatio-temporal resolution control. Novel, but simple, 3-D filter banks are proposed by using invertible deinterlacers which we have proposed before [8–10]. The proposed filter banks analyze a progressive video for yielding subband signals of a half frame rate in the progressive manner. Thus, a hierarchical structure is simply achieved. In addition, our filter banks also offer interlaced videos as intermediate data, where the VT Quincunx and FCO format are dealt with [7]. Furthermore, the reconstruction of the original full frame-rate progressive video is guaranteed by the PR property. The structure resembles to the lifting scheme, but unique in that the split process yields interlaced videos and joins with the deinterlacers.



(a) Line-based deinterlacing (b) Line-based reinterlacing

(c) Point-based deinterlacing (d) Point-based reinterlacing

Fig. 2: Examples of invertible deinterlacing and reinterlacing, where the symmetric extension method is applied. The white, black and gray circles denote pixels on the top, bottom and deinterlaced bottom fields, respectively.

2. REVIEW OF INVERTIBLE DEINTERLACING

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 [8–10]. This technique is different from a major deinterlacer [11] 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 1 summarizes the video formats before and after deinterlacing with sampling density preservation for both cases. This paper refers to the process shown in Fig. 1 (a) as *line-based deinterlacing* and (c) as *point-based deinterlacing* for the sake of consistent presentation. We also refer to a progressive video as '*frame sequence*' and to an interlaced video in either of the VT Quincunx or FCO format as '*field sequence*.'

Figure 2 (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. 2 (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. In the article [8], we have shown that there exist more general solutions and that this process can be regarded as a generalization of the simple field interleaving.

The invertible deinterlacing is not restricted to line-based field sequences and we can successfully apply this technique to point-based field sequences [8]. Figure 2 (c) illustrates an example of point-based invertible deinterlacing and (d) 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.

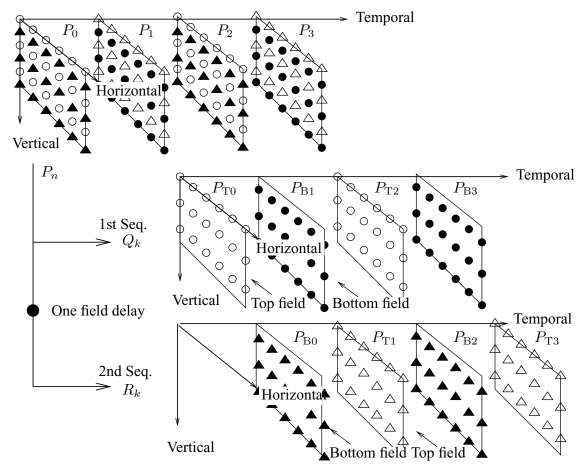


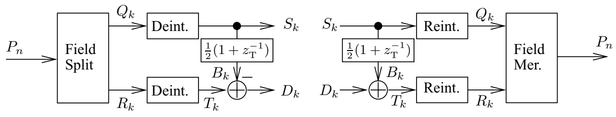
Fig. 3: Line-based field split process.

3. PROPOSED PR DEINTERLACER BANKS

For developing a novel spatio-temporal scalability, let us propose *perfect reconstruction (PR) deinterlacer banks* by using the invertible deinterlacers reviewed in the previous section.

3.1. Field Split Process

First of all, we suggest a line- and point-based field split process. Figure 3 shows the line-based case. In the same way, the point-based field split process is defined, although we omit to show it due to the limited space of paper. In these processes, input frame



(a) Deinterlacer bank (b) Reinterlacer bank

Fig. 4: Structures of proposed PR deinterlacer banks.

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.

3.2. Deinterlacer Banks

Combination of a field splitter, two invertible deinterlacers and a frame predictor yields a novel 3-D analysis filter bank as shown in Figure 4 (a), which we refer to as a *deinterlacer bank*. The deinterlacer bank 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. Frame sequence D_k is given as prediction errors so as to exploit the temporal redundancy, where a bi-directional frame prediction is shown as an example, that is $S_k = \mathrm{Deint}(Q_k)$ and $D_k = \mathrm{Deint}(R_k) - \frac{1}{2}(S_k + S_{k-1})$. Figure 4 (b) shows the inverse transform, namely *reinterlacer bank*, where 'Field Mer.' denotes field merge process, namely the inverse process of the field split.

The structures shown in Fig. 4 resemble to those of the 1/3 transform presented in the LIMAT framework [4, 5], and any motion compensation (MC) technique can be applied to the frame prediction. In the LIMAT framework, it is shown that adding an update lifting step, namely the 5/3 transform, improves the performance in terms of the coding efficiency. In our framework, however, the 1/3-transform-like structures shown in Fig. 4 have an advantage that receiving only the lower frame-rate progressive video S_k is sufficient for reconstruct the interlaced video Q_k

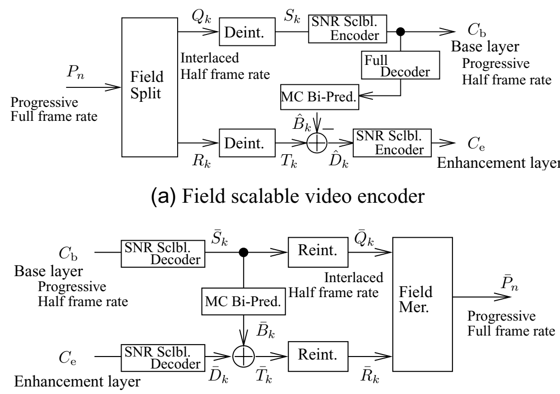
4. PERFORMANCE EVALUATION

Let us evaluate the performance of our deinterlacer banks. In this evaluation, we suggest a two-layer spatio-temporal scalable video codec system as a primal experimentation model.

4.1. Experimental model

Figure 5 shows the experimental video codec model of our proposed deinterlacer banks, which we refer to as a two-layer *field scalable video codec (FSVC)* system. It offers a scalable functionality that decoding base-layer stream C_{b} by itself provides both of half-rate frame sequence $\bar{S}_k$ and field sequence $\bar{Q}_k$, and adding enhancement-layer stream C_{e} improves the spatio-temporal resolution and yields full-rate frame sequence $\bar{P}_n$. In Fig. 5, 'SNR Sclbl. Encoder' and 'SNR Sclbl. Decoder' denote an SNR scalable encoder and decoder, respectively. In addition, 'MC Bi-Pred.' denotes a bi-directional frame predictor with MC.

For obtaining the experimental results in the followings, the deinterlacers shown in Fig. 2 were used and JPEG2000 was placed to the SNR scalable picture codec [6, 12], where fractional parts



(a) Field scalable video encoder

(b) Field scalable video decoder

Fig. 5: Primal experimentation model of a two-layer field scalable video codec (FSVC) system.

were maintained by scaling and using bit-depths greater than 8bpp. For the line-based case, a gain compensation technique was theoretically applied [10]. As for the point-based case, we experimentally found and applied energy gain factors 1.000 for LL1, 0.685 for LH1, 0.685 for HL1 and 0.469 for HH1. In addition, an MPEG-2-like half-pel full-search block matching technique was employed for the bi-directional prediction. To avoid the influence of quantization errors from the base layer to the enhancement layer, a local full-rate decoder, denoted as 'Full Decoder,' is employed. Note that the local decoder currently encounters some interesting problems such as decoder mismatch in the partial decoding case and feedback from deeper stages in the multi-stage case. With regard to these issues, further investigation is required.

4.2. Coding efficiency

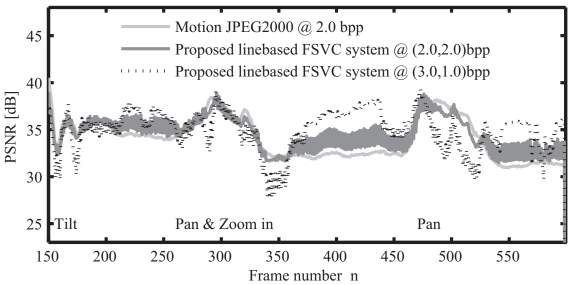
In order to verify the coding efficiency, we compare PSNRs between Motion JPEG2000 (MJP2) and the FSVC system. Figure 6 (a) and (b) respectively show the results of the line- and point-based system, where the luminance component of ITE[1] standard SIF video '*Whale Show*' (352×240P, 30Hz, 8-bit gray-scale) was used. In the graphs, $(2.0, 2.0)$- and $(3.0, 1.0)$-bpp describe bit-rate assignments, where $(r_{\mathrm{b}}, r_{\mathrm{e}})$-bpp implies that r_{b} and r_{e} bpp are assigned to the base and enhancement layer, respectively. Motion vectors are not counted in the results. Since MJP2 does not exploit the temporal redundancy, it is shown just for a reference.

From Fig. 6, it is noticed that the FSVC system moderately works for frames in little variation, whereas it suffers from degradation mainly due to camera activities, such as tilt, pan and zoom, especially in the point-based system. As well, it is shown that some dynamic bit-rate assignment procedure between layers is required. Our conjecture on the degradation of the point-based case is that remaining dot-pattern artifacts affect the quality and it could be improved through further investigation on the deinterlacers, MC techniques and the treatment of HH1 components in MJP2.

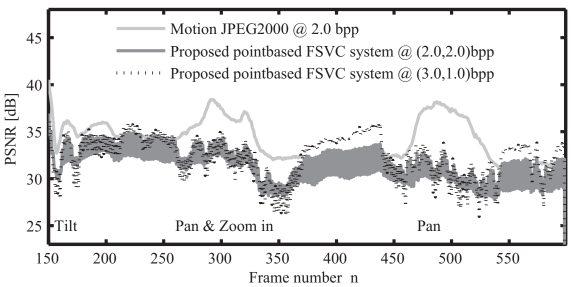
4.3. Suppression of comb-tooth and dot-pattern artifacts

Figure 7 shows four pictures obtained by decoding only the base-layer streams C_{b} at 0.2 bpp, where (a) and (c) are pictures through

[1] The Institute of Image Information and Television Engineers of Japan.



(a) Line-based FSVC system



(b) Point-based FSVC system

Fig. 6: Coding efficiency of the FSVC systems in PSNR of $\bar{P}_n$ encoded and decoded at 2.0 bpp in average (5.0688 Mbps) for the luminance component of '*Whale Show*' (SIF, 30P).

(a) Line-based interleaving

(b) Line-based deinterlacing

(c) Point-based interleaving

(d) Point-based deinterlacing

Fig. 7: Parts of half-rate frame pictures $\bar{S}_{75}$ (15P) decoded at 0.2 bpp (253.44 kbps).

the simple field interleaving as references, and (b) and (d) are results with the line- and point-based deinterlacer, respectively. From these results, the capability of deinterlacers for suppressing comb-tooth and dot-pattern artifacts can be verified. Summarizing, one can obtain low-bit-rate field sequences without the annoying artifacts through the simple deinterleaver instead of using the reinterlacer as well as frame sequences of a half frame rate.

5. CONCLUSION

In this work, we proposed novel 3-D filter banks, *i.e.* deinterlacer banks, which bring a new spatio-temporal scalability to video coding that partial decoding can yield both of progressive and interlaced videos of reduced frame rate by taking the SNR scalability into account. The VT Quincunx and FCO format were dealt with as interlaced video formats. Some primal experiments showed the potential of our proposal in practical applications. As future works, we will investigate the bit-assignment control, multi-stage structure [4, 5], variable-coefficient deinterlacers [9] and so on.

6. REFERENCES

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

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

[3] R. K. Rajendran, 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.

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

[5] 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.

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

[7] J. Kovačević and M. Vetterli, "FCO sampling of digital video using perfect reconstruction filter banks," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 118 –122, Jan. 1993.

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

[9] T. Ishida, S. Muramatsu, H. Kikuchi, and T. Kuge, "Invertible deinterlacing with variable coefficients and its lifting implementation," *IEEE Proc. ICASSP*, vol. 3, pp. 105–108, Apr. 2003, presented at IEEE ICME'03.

[10] T. Ishida, S. Muramatsu, and H. Kikuchi, "DWT gain compensation of Motion-JPEG2000 for invertible deinterlacer," *IEEE Proc. ICIP*, pp. 223 – 226, Sept. 2003.

[11] G. de Haan and E. B. Belleres, "Deinterlacing-an overview," *Proc. IEEE*, vol. 9, no. 86, pp. 1839–1857, Sept. 1998.

[12] Cannon, EPFL, and Ericsson, "http://jpeg2000.epfl.ch," URL.