IMPROVED SPATIAL SCALABILITY FOR VIDEO COMPRESSION

Vipan KAKKAR
School of Electronics & Comm. Engg. SMVD University,
Sub Post Office, Katra, vipan.kakar@smvdu.ac.in

Abstract: Spatial scalability (layering) in MPEG is not often used due to the lack of coding efficiency and available ICs. This work presents an alternative spatial scalability schemes that have better coding efficiency, and yet are relatively easy to implement with existing building blocks. Furthermore it presents experiments that have been conducted in order to improve the coding efficiency, and enable a backwards compatible HDTV DVD system. Attempts have been made to improve the spatial scalability of MPEG-2/MPEG-4.

Keywords: *Video Compression, MPEG, Motion Estimation, Spatial Scalability*

I. INTRODUCTION

Most video compression standards (e.g. MPEG-2, MPEG-4, H26L) support these scalability techniques. Spatial scalability is not often used due to the lack of coding efficiency. This paper presents some techniques for spatial scalability that do not have efficiency loss, yet are relatively easy to implement and have the potential of improved efficiency over known compression techniques. In this work spatial scalable schemes that generate streams that contain two resolutions (e.g. HD and SD) are presented. For a more detailed description of the MPEG-2 spatial scalability and its performance the reader is referred to [1]. Future applications will support a diverse range of display resolutions and transmission channel capacities, with the development of scalable extension to the state-of-the-art video coding standard [4]–[6]. This Scalable Video Coding provides support for multiple display resolutions within a single compressed bit stream (or in hierarchically related bit streams), which is referred to here as spatial scalability.

An example of an application using above mentioned spatial scalable schemes, is a DVD-disk with a bit stream that consists of two parts. One part of bit stream contains the MPEG-2 [1] normal (SD) resolution, whereas the other part of the bit stream contains an enhancement bit stream. The SD-bit stream is like a normal DVD stream, playable with regular DVD-players at the same quality. Extended DVD-players produce an HD resolution sequence using both, SD and enhancement part of the bit steams. These two parts of a bit stream can be stored on the DVD disk by either using multi-angle pointers or using one layer for the SD part and the other one for the enhancement part. This technique of spatial scalability is not restricted to DVD, it can for example also be applied for broad-cast, video over IP and wireless in home networks.

II. SPATIAL SCALABILITY

This section presents some derivatives of different schemes that improve the coding efficiency, yet are relative easy to implement. From each scheme the base layer will be a regular MPEG-2 encoder to be backward compatible with existing devices.

In Figure 1 a block diagram of MPEG-2/MPEG-4 spatial scalability is depicted. For the base layer a regular MPEG-2/MPEG-4 encoder can be used. Whereas for the enhancement layer a modified MPEG-2/MPEG-4 encoder has to be used. The enhancement encoder uses an upsampled version of base layer. For the upsampling poor interpolation filters are used. This is one of the main causes of the poor performance of this scheme. In the motion compensation loop of the enhancement layer encoder the upsampled base layer is added using a weighted criteria. This is another cause of its poor performance. The performance is poor because the coding efficiency of this layered coding scheme is not efficient. This means that, for a given picture quality, the bit rate of the base layer and the enhancement layer for a sequence together is substantially more than the bit rate for the same sequence coded at once, directly. An advantage of such a scheme, and of all scalable schemes is that it can be better resil-ience to transmission errors. This is the case when the base layer is transmitted via a channel with a low bit-error-rate (BER) and the enhancement layer is transmitted via a channel with a higher bit-error-rate.

II.1. KNOWN SCHEMES

The issue that the spatial scalability does not give a proper coding efficiency is a known fact. A block diagram of the scheme proposed by Dolby is depicted in Figure 2. The base-encoder can be a regular MPEG-2 encoder. The locally decoded image is upsampled using good upsampling filters. This upsampled image is than subtracted from the original

Manuscript received October 17, 2011; revised February 24, 2012

image resulting in a residue image. This residue image is than encoded using an MPEG-2 like encoder. To help the motion estimation of the enhancement encoder the scaled motion vectors of the base layer are used, indicated by the dashed line in Figure 2.

One of the advantages of this scheme over the MPEG-2/MPEG-4 scheme is that this scheme uses good upsampling filters.

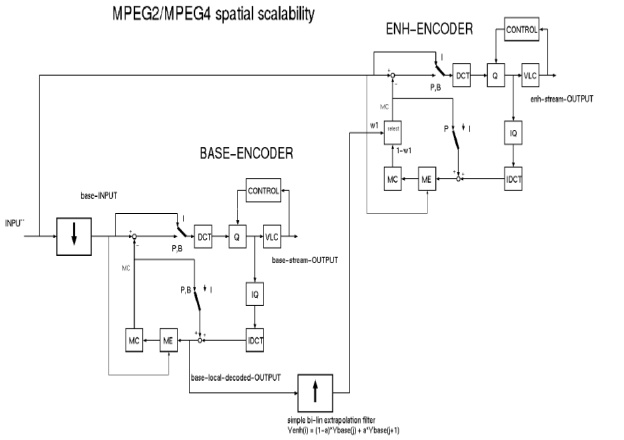

Figure 1: Block diagram of a merged decoder and encoder.

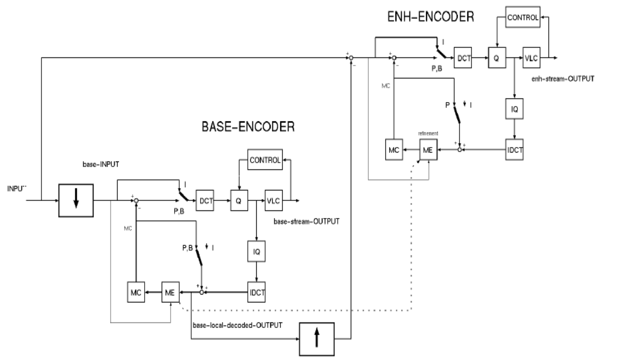

Figure 2: Block diagram of spatial scalability proposed by Dolby.

Such filters improve the coding efficiency significantly. An issue with this scheme is interlace sequences. It is also known that interlace is difficult to cope with properly. For this scheme hardly any investigations have been conducted into the interlace issue. For a more detailed description of this scheme the reader is referred to [2].

III. PROPOSED MOTION ESTIMATION SCHEME

In Figure 3, a block diagram of the Spatial Scalability with proposed Motion Estimation scheme is depicted. This scheme has quite some similarities with the scheme depicted in Figure 2. Two major differences have been proposed that are the position of the motion estimator and the DC offset.

Due to the fact that the motion estimation is done on the complete image instead of the residue, vectors are found that track the actual motion better. This leads especially for the consumer applications, hence lower bit rates than for the professional application, to a better picture quality.

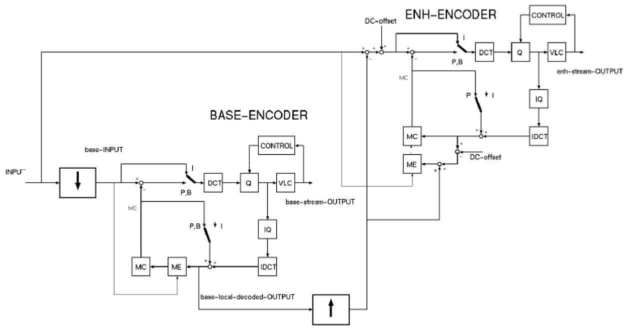

Figure 3: Block diagram spatial scalability with improved motion estimation.

Furthermore a DC-offset is introduced with this offset a DC-value can be added to the residue signal. The residue signal is normally concentrated around zero. By adding a DC-offset the samples can be shifted to the middle of the range (128 for 8 bit video samples).

The advantage of this addition is that standard building blocks for the encoder of the enhancement layer can be applied resulting in a cost efficient (re-use of IP-blocks) solution.

A block-diagram of a decoder of such a spatial scalable scheme is depicted in Figure 4. It contains two regular MPEG-2 decoders. The base decoder operates at the lower resolution, e.g. SD, whereas the enhancement decoder works on the full resolution. The output of the base decoder can be viewed on a regular TV-set. The output of this decoder is upscaled to HD resolution and this upscaled signal is fed to the addition point. The signal of the decoded HD enhancement signal is fed to the addition point as well.

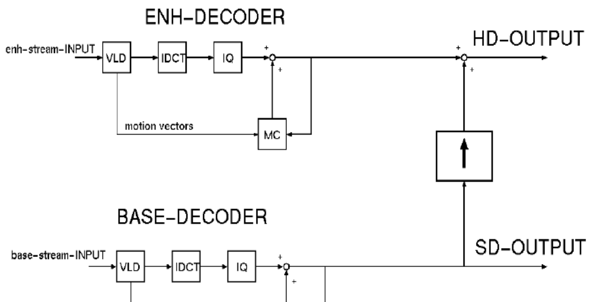

Figure 4: Block diagram spatial scalability decoder.

Both signals are then added and the DC-offset will be subtracted from the resulting signal. The resulting signal is now the complete HD signal.

III.1 CODING EFFICIENCY IMPROVEMENTS

This section deals with optimisations to the spatial scalable coding schemes. These optimisations are discussed and their impact on the implementation, if applicable. Further-more the results obtained with these optimisations will be presented, especially for these sequences it is important to have a look at the sequences as well. Due to the layered approach SNRs do not make much sense and are therefore left out.

IV. EXPERIMENT SET-UP

In this comparison non-scalable streams with scalable streams are compared. That means that a sequence is directly coded to a particular reasonable bit rate and than code that same sequence into a base layer stream and an enhancement layer stream. The bit rate of the two streams, base layer and enhancement layer stream, add up to more or less the same bit rate as the bit rate of the non-scalable stream. For example a difficult HD sequences are coded to 9.5 Mbps and coded that same sequence to an SD bit stream of 5.5 Mbps and an enhancement bit streams of 4 Mbps, adding up to 9.5 Mbps as well. These bit rates are typically required to be able to store one complete HD movie on a single double layer DVD. The ratio between the base layer and the enhancement layer are chosen such that the base layer could fit on one layer and the enhancement layer on the other layer although the formatting does not have to be like that. This ratio in general makes sense to more or less evenly distribute between the two bit streams.

IV.1 INITIAL EXPERIMENTS

The initial experiments with spatial scalability were done using the schemes presented in Section 1, (MPEG-2/MPEG-4 spatial scalability, Dolby spatial scalability). The results are depicted in Table 1. The bit rates for these two schemes are somewhat higher than the non-scalable stream. The non scalable stream has the best quality although the quality of the Dolby scheme comes close to the quality of the non-scalable stream. The quality of the MPEG-2 spatial scalable scheme is from a perceptual point of view worse than the other two methods.

One of the goals is to be as much as possible compatible with the MPEG-2 standard so the modifications to the existing hardware IP-block would be as small as possible.

	Non-scalable	Base Layer	Mpeg2	Improved Mpeg2	Proposed Scheme (IME)
Bitrate (Mbps)	9.52	5.5	4.77	4.5	4.1
Sum			10.28	9.99	9.6
Dev. Non-scalable (%)			7.55	4.91	3.7

Table 1: Spatial Scalability Results using Existing Techniques

Therefore the initial experiments were done using ½ pel motion compensation. This ½ pel motion compensation is the standard within MPEG-2. Single modifications at a time will be applied to evaluate their individual merit. So after each test the modification will be switched off again, return to the starting point. So the first experiments were performed using ½ pel motion estimation and adaptive quantization activated. The results with these settings are presented in Table 2 indicated with label "using ½ pel MC".

IV.2 QUARTER PEL MOTION COMPENSATION

In MPEG-2 the motion compensation has a ½-pel precision. This means that the motion vectors will have displacements between pixels. In case these fractional vectors are used adjacent pixels are bilinear interpolated. As described in, for example [5], this ½-pel precision is not optimal. A precision of a ¼-pel is considered to be nearly optimal. Furthermore using bilinear interpolation, using a [1 2 1]-filter is not optimal, better interpolation filters, e.g. [-1 0 9 16 9 0 –1], can be applied. On the other hand the vector cost will increase due to the ¼-pel accuracy.

		Non-scalable	Base Layer	Proposed Scheme
Using ½ pel MC	Bitrates (Mbps)	9.59	5.55	4.30
	Sum			9.58
	Dev. Non-scalable (%)			3.6
Using ¼ pel MC	Bitrates (Mbps)	9.59	5.55	4.05
	Sum			9.6
	Dev. Non-scalable (%)			1.01

Table 2: Spatial Scalability Results using Proposed Scheme

In Table 2 is depicted that usage of ¼-pel motion compensation with better interpolation filters, labelled "using ¼ pel MC", gives some decrease in bit rate, about 5 % for the enhancement layer only and 2.5 % for the overall bit rate. Together with the reduction in bit rate the quality slightly improves, slight improvements can be observed at the edges of objects.

IV.3 H26L FOR ENHANCEMENT LAYER CODING

H26L is known to have a better coding efficiency than MPEG-2 especially on the lower bit rates. Therefore, we investigated the coding gain on the enhancement layer using H26L, for the base layer we stick to MPEG-2 in order to stay backward compatible. For a detailed description of the H26L standard the reader is referred to [7].

The results using H26L are depicted in Table 3. Using H26L for the enhancement layer reduces the bit rate substantially, whereas the picture quality is comparable. The bit rate of the enhancement layer is reduced by more than a factor of two.

	Non-scalable	Base Layer	H26L
Bitrates (Mbps)	9.59	5.55	1.8
Sum			7.35
Dev. Non-scalable (%)			-22

Table 3: Spatial scalability results using H26L

V. CONCLUSIONS

It has been seen that spatial scalability can be performed without sacrificing much coding efficiency compared to direct encoding. Whilst the implementation of this spatial scalability remains relatively easy. Furthermore, the work presents a scheme that allows for gradual tradeoffs between bit rate and picture quality.

The presented techniques can be an enabler for many new applications. They can be found in the area of wireless in-home video distribution and video over IP (QOS). Or in broadcast type of applications where compatibility can be an important consideration. For other applications the possibility of a 'in the clear' base stream and scrambled enhancement stream would be the decisive reason. Also distributed storage applications could benefit from these techniques.

REFERENCES

[1] B. Haskell , A. Puri and A. N. Netravali Digital video: An introduction to MPEG-2. :Chapman & Hall, 1997

[2] G. Demos, Temporal and resolution layering in advanced television, DemoGraFX (Dolby) SPIE Proceedings Vol. 2663, pp 52-68

[3] B. Girod Motion Compensation Prediction with Fractional-Pel Accuracy, IEEE Transactions Communications, Vol 41, 604-12, April 1993.

[4] G. J. Sullivan and T. Wiegand "Video compression-from concepts to the H.264/AVC standard", Proc. IEEE, vol. 93, p.18 , 2005.

[5] H. Schwarz, D. Marpe and T. Wiegand "Overview of the scalable video coding extension of the H.264/AVC standard", IEEE Trans. Circuits Syst. Video Technol., vol. 17, p.1103 , 2007.

[6] D. Marpe , T. Wiegand and G. J. Sullivan "The H.264/MPEG4 advanced video coding standard and its applications", IEEE Commun. Mag., vol. 44, p.134 , 2006.

[7] ITU Telecommunication Standardization, Video Coding Expert Group, http://standards.pictel.com/ ftp/video-site/h26L