SNR Scalability with Matching Pursuits

C. De Vleeschouwer*and B. Macq‡

Abstract

In this paper, SNR scalable representations of video signals are studied. The investigated codecs are well suited for communications applications because they are all based on backward motion-compensated predictive coding, which provides the necessary low-delay property. In a Very-Low Bit Rate context (VLBR), the Matching Pursuits signal representation algorithm is used to represent the Displaced Frame Difference (DFD) of each layer of a multilevel decomposition of the video signal. A number of conventional prediction schemes that can be generalized to any DFD representation technique are considered. They are compared with an original and MP specific DFD prediction method.

1 INTRODUCTION

In addition to the new requirements for real-time transmission, video applications are quickly evolving from one-to-one communications to one-to-many and many-to-many communications. Widespread use of these applications can easily overload existing networks when the same bits of information have to be transmitted to different users at the same time. Multicast, where a single packet is addressed to all intended recipients and the network replicates packets only as needed, make it possible to reduce unnecessary duplication of bits and relieve some of the network load [1, 2]. In this context, a single, fixed-rate stream can not satisfy the conflicting requirements of a heterogeneous set of receivers. One approach for delivering multiple resolution, frame rate, or levels of quality across multiple network connections is to encode the video signal with a set of independent encoders, each producing a different output rate. This approach, called **simulcast**, does not exploit statistical correlations across subflows. Its compression performance is thus suboptimal. By contrast, a layered, or scalable, coder exploits correlations across subflows to achieve better overall compression [3, 4]. The input signal is compressed into a number of discrete layers, arranged in hierarchically to provide progressive refinement. With *scalable* coding algorithms, one original compressed video bitstream is generated; but different subsets of the bitstream can then be selected at the decoder end to support a multitude of display specifications such as the quality level (SNR scalability), the spatial resolution (spatial scalability), and the frame rate (temporal scalability).

Developing (spatially) scalable video compression algorithms has attracted considerable attention in recent year [5, 6, 7]. In this paper, we focus our attention to the study of SNR scalable schemes based on the extension of the hybrid motion prediction/Matching Pursuits coding algorithm [8].

Section 2 discusses the problem encountered when coping with SNR scalability. In section 3, a method is proposed to predict the high SNR layer DFD from the low SNR layer reconstructed DFD. This method is specific to the use of the Matching Pursuits (MP) representation technique

*The research of C. De Vleeschouwer is supported by the Belgian N.F.S.

†Corresponding author.

‡Part of this work has been submitted for publication in IEEE Transactions on Multimedia.

and has been implemented to minimize the size of either the low SNR layer bitstream, or the complete scalable bitstream. For the sake of comparison, other prediction schemes have been considered in section 4. They include frame-based or macroblock-based DFD prediction modes but also propose others. Results and discussions are provided in section 5.

2 SNR SCALABILITY: TERMS OF THE PROBLEM

As told before, SNR scalability allows for the decoding of appropriate subsets of a single output bitstream to generate gradual quality approximations of the original sequence. In the following, despite its generalization to multi-level decompositions is straightforward, studies are restricted to a two-layer system. The **low SNR layer** contains coded information required for low quality delivery. The **high SNR layer** provides additional information required to display a high quality sequence. We refer to the frames generated using both the low and high SNR layers information as to high SNR layer frames.

Given a quality constraint for each layer, the designer of the video scalable coder has to choose between two distinct objectives, depending on the application (s)he deals with.

1 **Low quality delivery with minimal bitrate:** the purpose is to minimize the size of the bitstream subset providing the low quality video display.

2 **High quality delivery with minimal bitrate:** the purpose is to minimize the total size of the bitstream, i.e. the size required to achieve high visual quality display.

For example, for multicast video transmission, depending on its goal (wide distribution of advertising, delivering of a high quality service to the users who pay while providing a minimal service to others, ...), the source may desire to favor the high or low SNR layer at the expense of the other. In the next sections, different strategies are investigated to allow SNR scalability for Matching Pursuits (MP) video coding.

3 ATOM-BASED PREDICTION USING MATCHING PURSUITS

The Matching Pursuits video coding algorithm is a hybrid motion-compensated algorithm for which the Displaced Frame Difference (DFD) is expanded into waveforms, called *atoms*, chosen among an overcomplete dictionary [8, 9]. Stating that a motion-compensation loop is used in each layer, the aim of this section is to study and exploit the statistical dependencies between the prediction errors or Displaced Frame Difference (DFD) of both layers to design an efficient scalable codec. The ability of Matching Pursuits to perform a signal analysis when decomposing it into waveforms is used to predict the main structures of the high SNR layer DFD from the low SNR layer reconstructed DFD.

3.1 Layers Prediction Errors (DFD)

Figure 1 presents the components of each layer of the scalable scheme. $I_o(t-1)$ and $I_o(t)$ refers to the original picture at time $t-1$ and t respectively. Given a **single** set of motion vectors $V(t)$ and the linear motion-compensation operator $\Gamma(.,V(t))$, the motion-compensation provides $P_o(t) = \Gamma(I_o(t-1), V(t))$. We can observe that the frame difference $DFD_o(t) = I_o(t) - P_o(t)$ is not strictly null due to the fact that motion compensation is not able to perform perfect prediction (non-linearity of the motion, occlusions, new objects, ...). Actually, this

frame difference is not computed in a predictive scheme because, for convergence purpose, it is necessary to compute the difference between the original frame and the previously reconstructed compensated frame. $I_l(t-1)$ and $I_h(t-1)$ being the low and high SNR reconstructed frames at time $t-1$, the motion-compensation provides a prediction at time t for the low and high SNR layers, denoted $P_l(t)$ and $P_h(t)$ respectively. The prediction error (DFD) is then computed for each layer:

$$\begin{aligned} DFD_l(t) &= I_o(t) - P_l(t) \\ &= I_o(t) - \Gamma(I_l(t-1), V(t)) \\ DFD_h(t) &= I_o(t) - P_h(t) \\ &= I_o(t) - \Gamma(I_h(t-1), V(t)) \end{aligned} \quad (1)$$

Denoting $E_l(t-1) = I_o(t-1) - I_l(t-1)$ and $E_h(t-1) = I_o(t-1) - I_h(t-1)$ the reconstruction (coding) error of the low and high SNR layers at time $t-1$, we obtain that:

$$\begin{aligned} DFD_l(t) &= I_o(t) - \Gamma(I_l(t-1), V(t)) \\ &= I_o(t) - \Gamma(I_o(t-1) - E_l(t-1), V(t)) \\ &= DFD_o(t) + \Gamma(E_l(t-1), V(t)) \end{aligned} \quad (2)$$

$$\begin{aligned} DFD_h(t) &= I_o(t) - \Gamma(I_h(t-1), V(t)) \\ &= I_o(t) - \Gamma(I_o(t-1) - E_h(t-1), V(t)) \\ &= DFD_o(t) + \Gamma(E_h(t-1), V(t)) \end{aligned} \quad (3)$$

So, linearity of the motion-compensation operator allows expressing both low and high SNR prediction errors as the sum of two terms. The first one, $DFD_o(t)$, is common for both layers. It is due to the erroneous prediction obtained when applying motion compensation to the original picture. The second one results from the introduction of the residual coding error within the prediction loop. This term differs in both layers.

The presence of a common term in the prediction errors of both layers allows the encoded/decoded low SNR layer DFD to predict the high SNR layer prediction error. Matching Pursuits, by their ability to analyse locally the signal when representing it, allow selecting predictable structures among both errors. Two selection strategies are proposed in sections 3.2 and 3.3.

3.2 Low SNR Layer Delivery with Minimal Bitrate

Delivering the low quality with a minimum bitrate means that the non-scalable video coder generates the low SNR layer bitstream. In order to provide scalability, additional information must complete it in order to achieve a better quality when the complete information is accessed.

Following the prediction structure of figure 2 and given the **motion vectors set $V_l(t)$ used to predict the low SNR layer** at time t, one may compute the high SNR layer prediction $P_h(t) = \Gamma(I_h(t-1), V_l(t))$. The resulting prediction error, i.e. $DFD_h(t)$, has to be encoded using Matching Pursuits. Nevertheless, as the low and high SNR prediction errors have common structures, it is highly probable that atoms used to represent the low SNR layer prediction error also match the high SNR layer DFD structures.

The high SNR layer DFD representation algorithm is modified as follows. At each step of the algorithm, i.e. each time an atom has to be selected, two possibilities are considered:

- as in the conventional MP representation technique, the best matching dictionary function is searched for. Note that this search indirectly selects the (Y,U,V) component, denoted C_{YUV}, of the atom.

- among the functions defining the low SNR layer atoms (shape + position + (Y,U,V) component), the best matching function is also searched for. Note that only the atoms of the low SNR layer belonging to the C_{YUV} (Y,U,V) component are considered.

In order to decide which one of the two pre-selected atoms is really transmitted, a Rate/Distortion criterion is used. In the first case, the atom is specific to the high SNR layer. Its coding cost can be estimated from the coding cost of the atoms that have been encoded in the previous frame. We denote $COST_{h,\bar{l}}$ the mean coding cost of the high SNR layer atoms that do not belong to the low SNR layer. In the second case, the atom has already been defined in the low SNR layer. Only its amplitude has to be transmitted. In practice, a VLC code transmits it differentially with respect to the amplitude in the low SNR layer (and an escape code is used for the atoms of the low SNR layer that are not used in the high SNR layer). The coding cost is estimated by the mean differential coding cost of the atoms of the previous frame that are encoded in the same way. We denote it $COST_{h,l}$. For both atoms, the decrease of distortion achieved by the representation of the atom is the atom energy, i.e. the square of its amplitude, respectively $A_{h,\bar{l}}$ and $A_{h,l}$. In order to maximize the ratio $-\Delta D/\Delta R$, the atom specific to the high SNR layer is selected if:

$$\frac{A_{h,l}^2}{COST_{h,l}} \leq \frac{A_{h,\bar{l}}^2}{COST_{h,\bar{l}}} \tag{4}$$

3.3 High SNR Layer Delivery with Minimal Bitrate

The optimal way to provide the high quality level is to use the non-scalable MP codec for the high SNR layer. This non-scalable codec is used as reference. The aim is to generate a bitstream of minimal size providing a high quality sequence while allowing the extraction of a subset for the reconstruction of a low quality sequence.

Figure 3 presents the strategy followed to reach our aim. The high SNR layer is used as reference to estimate the set of motion vectors. Following the conventional non-scalable scheme, atoms are selected to represent the high SNR layer prediction error, i.e. $DFD_h(t)$. Once this set has been selected, the cheapest way to provide a lower quality layer is investigated, assuming that prediction is possible between atoms of the low SNR layer and atoms of the high SNR layer.

First, **the low SNR layer prediction is computed using the motion vectors estimated on the high SNR layer**, i.e. $P_l(t) = \Gamma(I_l(t-1), V_h(t))$. The prediction error is represented using either an atom specific to the low SNR layer, or one that has been defined for the high SNR layer. R/D considerations lead to the optimal choice. $COST_l$ is the mean coding cost of defining a low SNR layer atom in the previous frame, $COST_{h,\bar{l}}$ is the mean cost of defining an high SNR layer atom that does not belong to the low SNR layer, and $COST_{h,l}$ is the mean cost of transmitting the differential amplitude between the two layers for the atoms that belong to both layers in the previous frame. $A_{l,h}$ is the amplitude, in the low SNR layer, of the atom that had been preselected in the high SNR layer. $A_{l,\bar{h}}$ is the amplitude of the atom specific to the low SNR layer. In order to maximize the ratio $-\Delta D/\Delta R$, the atom specific to the low SNR layer is selected if:

$$\frac{A_{l,h}^2}{COST_l - COST_{h,\bar{l}} + COST_{h,l}} \leq \frac{A_{l,\bar{h}}^2}{COST_l} \tag{5}$$

Once enough atoms have been selected in the low SNR layer to achieve the required quality, the vectors $V_h(t)$ and the atoms of the low SNR layer are encoded and the low SNR layer bitstream is generated. To provide the high quality bitstream, it is completed by the VLC

codes defining the differential amplitude of the atoms present in both layers (conditional entropy coding) and by encoding the features of the remaining atoms of the high SNR layer.

4 BLOCK- or FRAME-BASED PREDICTION

For the sake of comparison and before the presentation of the results achieved with the atom-based prediction scheme, three other scalable schemes are considered. It is worth noting that, on the contrary of the atom-based prediction schemes, these three schemes are not specific to the use of the Matching Pursuits prediction error representation. They can be generalized to any hybrid motion-compensated scheme, whatever the DFD representation technique is. Some of the conclusions drawn from the simulations can thus be generalized to hybrid coders involving other DFD representation techniques.

4.1 MPEG-4 Scheme for Scalability using DCT

The first one is the versatile MPEG-4 DCT scalable scheme. In this scheme, the high SNR layer is bidirectionally predicted on a macroblock basis. The two references are the previous high SNR layer frame (with motion-compensation) and the current low SNR layer frame. Figure 4 presents the outline of the scheme. Solid lines represent the two prediction mode considered. When motion prediction is selected, an high SNR layer motion vector is encoded. **Two sets of motion vectors** (high and low SNR layer) are thus defined in this scheme.

4.2 MacroBlock-Based Prediction using Matching Pursuits

The second one is also a macroblock-based prediction scheme. It differentiates itself from the MPEG-4 scheme by the fact that the DFD is encoded using MP and by the fact that three modes of prediction are considered:

- MB estimation results from the motion compensation of the previous frame of the high SNR layer (MC mode),
- MB prediction is the sum of the motion compensation and of the low SNR layer reconstructed prediction error (MC+DFD mode),
- MB is predicted by the current low SNR layer reconstructed image (L-SNR-I mode).

The usefulness of a three-modes prediction strategy towards a conventional bimodal strategy (like the one used in the MPEG-4 scalable coder) is demonstrated in section 5.1.

4.3 Frame-Based Prediction of the DFD

The third one is the one proposed by UC Berkeley [9]. To generate the low SNR layer bitstream, a normalized weighted sum of both prediction errors is encoded, i.e. $(1-\gamma).DFD_l + \gamma.DFD_h$, $0 \leq \gamma \leq 1$. The reconstructed prediction error is added to the low SNR layer prediction to generate the low SNR layer frame. It is also added to the high SNR layer motion prediction to generate a new high SNR layer prediction, used as a reference for the high SNR layer generation. In [9], γ parameter is used to adjust the quality between both layers while keeping their relative bit rates fixed. Nevertheless, it appears that modifying the γ value only allow for a small improvement of the high SNR layer (between 0.1 and 0.5 dB) in return for a large quality degradation of the low SNR layer (between 2 and 3 dB). In section 5, it appears that estimating the set of motion vectors from the high SNR layer rather than from the low SNR one

allows for a larger decrease of the complete bitstream size than the one allowed by increasing the γ value.

5 RESULTS

All the presented results have been generated using a set of four video sequences selected among the ones recommended in the MPEG-4 video coding group for the test of VLBR coding algorithm. For the sake of comparison, common coding conditions (see table 1) have been fixed, whatever the scalable scheme is. PSNR quality levels (in dB) have been settled for both the low and the high SNR layers. Matching Pursuits allow respecting this constraint stringently as, for each frame, atoms can be added to the reconstructed prediction error until the required quality level has been reached. Comparison between the considered coding schemes is based on the size of the bitstreams required to achieve the PSNR quality constraint.

5.1 Low quality delivery with minimal bitstream

As the low SNR layer quality has to be provided with a minimal bitbudget, it is encoded like in the non-scalable scheme. In table 2, the bit-rate (in bits/s) is presented for five schemes. In the simulcast scheme, both layers are encoded independently. It is a non-scalable scheme proposed as a reference. The atom-based prediction scheme refers to the one presented in section 3.2, while the block-based and frame-based schemes refer to sections 4.2 and 4.3. Results obtained when using the MPEG-4 DCT-VM are also presented. In this case, the quantization parameter has been adjusted to provide a mean SNR quality that is just below the quality required in table 1. For each scheme, the bit-rates for transmitting only the low SNR layer, only the high SNR layer and both layers together are given. Of course, for scalable schemes, the bitstream providing the high SNR layer also provides the low SNR layer. On the contrary, for simulcast transmission, providing both the low and the high SNR layers requires both bitstreams' transmission.

An obvious conclusion that can be drawn from table 2 is the superiority of the MP schemes vis-a-vis the DCT MPEG-4 verification model scheme. Another conclusion is that all scalable schemes considered manifest very similar behavior-patterns. For the chosen test conditions, they allow saving about 20% of the total (low+high SNR layers) bitbudget used when providing the high and low quality level separately. It is obtained in exchange for an increase (15-25%) in the necessary bitbudget for providing the high SNR layer only.

Among the considered MP scalable schemes, the frame-based prediction is the most favorable from a computational point of view. Moreover, its performances often equal or outperform those of the other schemes. This prediction scheme can be recommended when trying to improve the quality of a non-scalable low SNR layer.

Nevertheless, for some sequences, the MacroBlock(MB)-based prediction scheme's performance is better than the frame-based one. Contrary to the frame-based scheme, which systematically predicts the high SNR layer DFD, the MB-based scheme allows either no prediction at all, or image MB prediction. These modes of prediction are useful for MBs where prediction errors differ in both layers. Referring to the discussion in section 3.1, these MBs are the ones for which the compensation of the previous frame's residual error differs for both layers and for which this difference is significant compared with the original image motion compensation error. This occurs when the translational motion model fits the motion of objects that are encoded with different qualities in both layers. The usefulness of 3 prediction modes is emphasized by Table 3. Using 3 modes always outperforms a bidirectional prediction. yet, when bidirectional prediction is used, prediction of the high SNR layer DFD is preferable to the prediction of the high SNR layer frame. For the "Coastguard" sequence, both modes are nearly equivalent. Ta-

ble 3 also presents the results that are obtained when using two distinct sets of motion vectors for the low and the high SNR layers. MBs that are Inter in the low SNR layer may either be predicted from the low SNR layer or motion-compensated using a motion vector specific to the high SNR layer. Due to the use of a set of optimal motion vectors to predict the high SNR layer, the motion prediction error of the high SNR layer has less energy. Nevertheless, it also suffers from a smaller correlation with the low SNR layer DFD and the efficiency of the prediction between the layers decreases. We have observed that the number of blocks predicted by the MC+DFD mode decreases when two sets of motion vectors are used. From table 3, it appears that the use of two sets of motion vectors is not useful. The additional motion vectors coding cost and the loss of correlation between the DFD structures of both layers outstrip the gain resulting from the improved motion prediction. It is worth noting that this improved motion prediction also requires a significant additional computational load.

5.2 High quality delivery with a minimal bitstream

Our aim is to investigate how to reduce the size of the complete scalable bitstream, i.e. how to generate a bitstream of minimal size permitting the delivery of two quality levels. Table 4 presents the bit-rates (bits/s) obtained for a number of scalable schemes. For all considered scalable schemes, the set of motion vectors has been estimated from the high SNR layer. It permits a significant reduction of the scalable bitstream size in return for a small increase in the low SNR layer bitbudget. This is observed when comparing the frame-based prediction ($\gamma = 0$) and the macroblock-based of table 4 with the ones of table 2. For the frame-based prediction scheme, a set of parameters has been considered. Most often, only small decreases in the complete scalable bitstream are obtained in return for large increases in the low SNR layer bitbudget. When trying to minimize the size of the complete bitstream, the best result is obtained with the atom-based prediction method presented in section 3.3. The bitbudget surplus of the scalable scheme in comparison with the non-scalable one is limited to 5-7%, which is a significant reduction with regards to the 20% surplus obtained when low quality was delivered with a minimal bitbudget. This is also a significant decrease (3-12%) compared to the surplus of all other frame-based and macroblock-based prediction schemes considered. As observed in table 4, the surplus reduction is obtained in exchange for a significant increase of the low SNR layer bit-rate. As told in section 2, the choice of the scalable prediction scheme should thus be application-driven.

6 CONCLUSIONS

A number of SNR scalable schemes based on the hybrid motion-compensated Matching Pursuits video coding algorithm have been considered. Given an SNR quality constraint for each layer, two distinct objectives of scalable codecs have been highlighted:

- to generate a scalable bitstream delivering the low quality (low SNR layer) with a minimal bit number,
- to generate a scalable bitstream with a minimal total bit number.

Comparisons between different prediction schemes between the low and high SNR layers have shown that the prediction efficiency of all schemes is more or less equivalent for the first aim. Computational or implementation complexity should guide the choice of the scalable codec designer. On the contrary, the achievement of the second objective is open to compromise and the choice of the prediction scheme should be application-driven. Based on the Matching

Pursuits DFD representation, an original atom prediction scheme allows for high compaction of the scalable bitstream in return for a significant increase in the low SNR layer bitbudget. Other conventional prediction schemes, that can be generalized to any DFD representation technique, only reach a limited compaction of the scalable bitstream. For these schemes, the estimation of the set of motion vectors on the high SNR layer brings a higher compaction of the total bitstream than the "heuristic" modification of the low SNR layer DFD coding strategy proposed in [9] ($\gamma \neq 0$).

Another lesson from this work is that the use of two sets of motion vectors is not useful. The additional motion vectors coding cost and the loss of correlation between the DFD structures of both layers outstrip the gain resulting from the improved motion prediction.

Efficiency of Matching Pursuits SNR scalable schemes has been demonstrated in comparison with the MPEG-4 generalized DCT scalable scheme. Robustness of MP scalable schemes towards transmission errors should be thoroughly investigated in future research.

References

[1] Dave Kosiur. *IP MULTICASTING: The Complete Guide to Interactive Corporate Networks.* Wiley Computer Publishing, New York, 1998. ISBN: 0-471-24359-0.

[2] C. Diot, W. Dabbous, and J. Crowcroft. Multipoint Communication: A Survey of Protocols, Functions, and Mechanisms. *IEEE Journal on Selected Areas in Communications*, 15(3):277–290, April 1997.

[3] N.F. Maxemchuk. Video Distribution on Multicast Networks. *IEEE Journal on Selected Areas in Communications*, 15(3):357–372, April 1997.

[4] S. McCanne, M. Vetterli, and Van Jacobson. Low-Complexity Video Coding for Receiver-Driven Layered Multicast. *IEEE Journal on Selected Areas in Communications*, 15(6):983–1001, August 1997.

[5] Klaus Illgner and Frank Muller. Spatially Scalable Video Compression Employing Resolution Pyramids. *IEEE Journal on Selected Areas in Communications*, 15(9):1688–1703, December 1997.

[6] D. Taubman and A. Zakhor. Multirate 3-D Subband Coding of Video. *IEEE Transactions on Image Processing*, 3(5):572–588, September 1994.

[7] M. Ghanbari and V. Seferidis. Efficient H.261-Based Two-Layer Video Codecs for ATM Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(2):171–175, April 1995.

[8] R. Neff and A. Zakhor. Very Low Bit Rate Video Coding Based on Matching Pursuits. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1):158–171, February 1997.

[9] O. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor. Video Compression Using Matching Pursuits. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):123–143, February 1999.

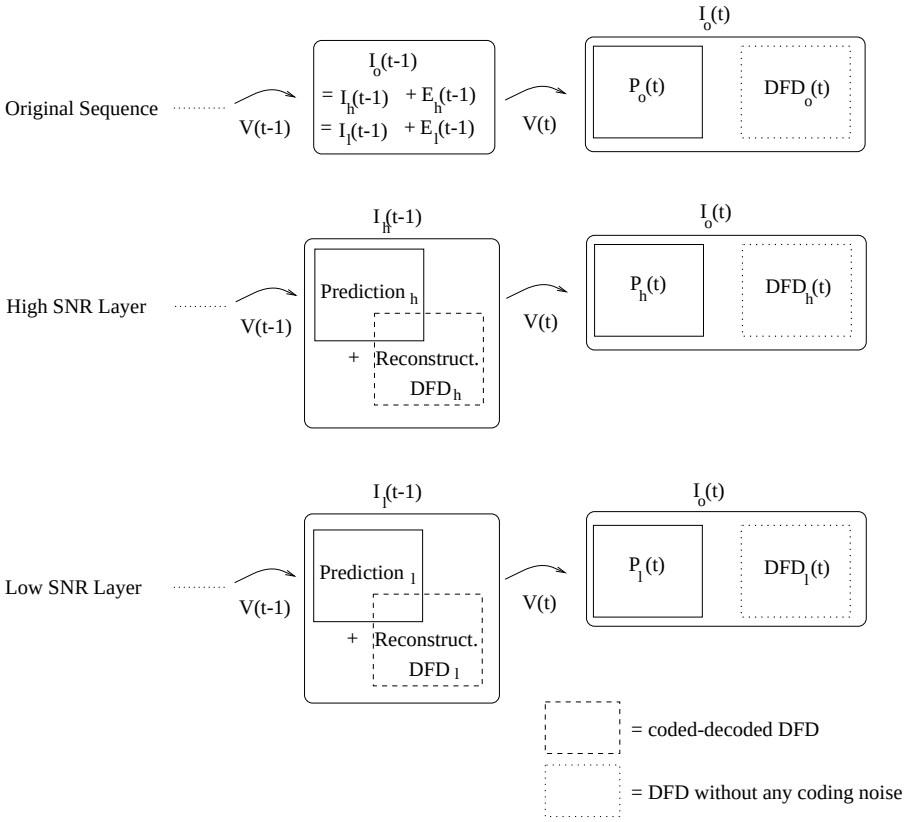

Figure 1: High and low PSNR layers illustration.

	Frame resolution	Frame rate	Base layer PSNR quality	Enh. layer PSNR quality
Hall	QCIF	7.5 f/s	31 dB	34 dB
Silent	QCIF	10 f/s	31 dB	34 dB
Foreman	QCIF	10 f/s	31 dB	33 dB
Coastguard	QCIF	10 f/s	30 dB	32 dB

Table 1: Summary of the coding conditions for each considered sequence.

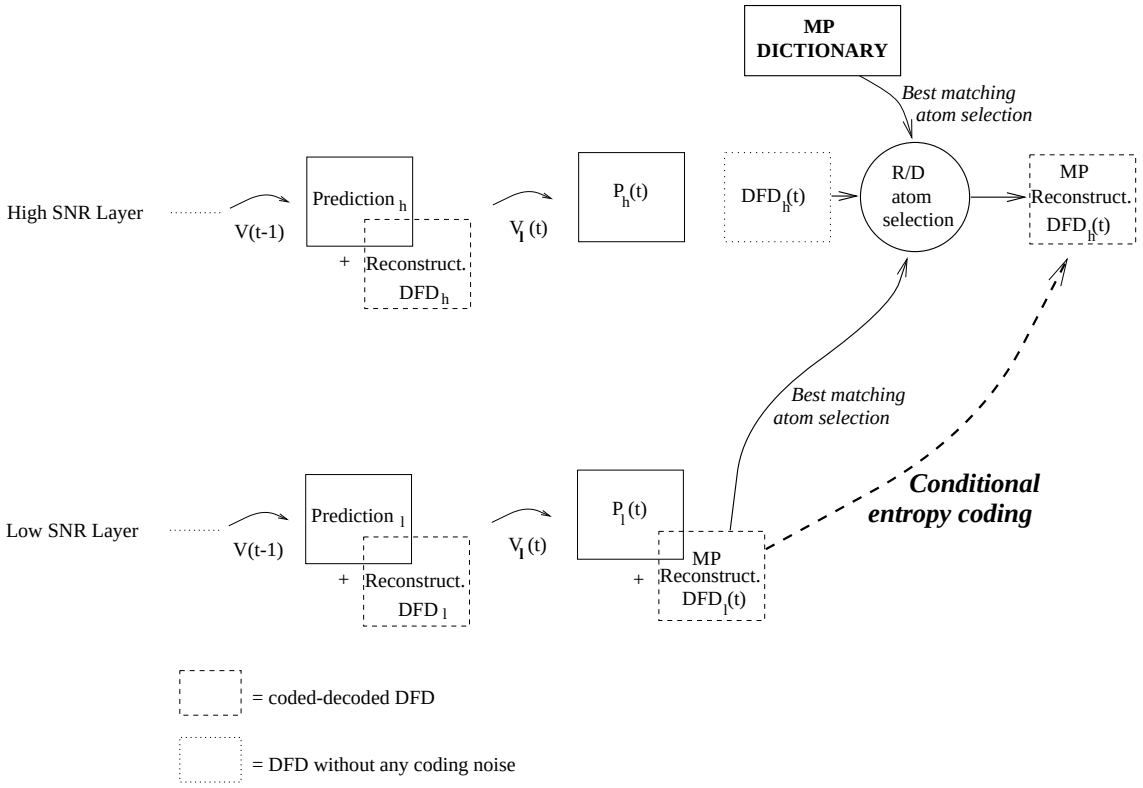

Figure 2: Atom-based prediction of the high SNR layer DFD from the low SNR layer reconstructed DFD.

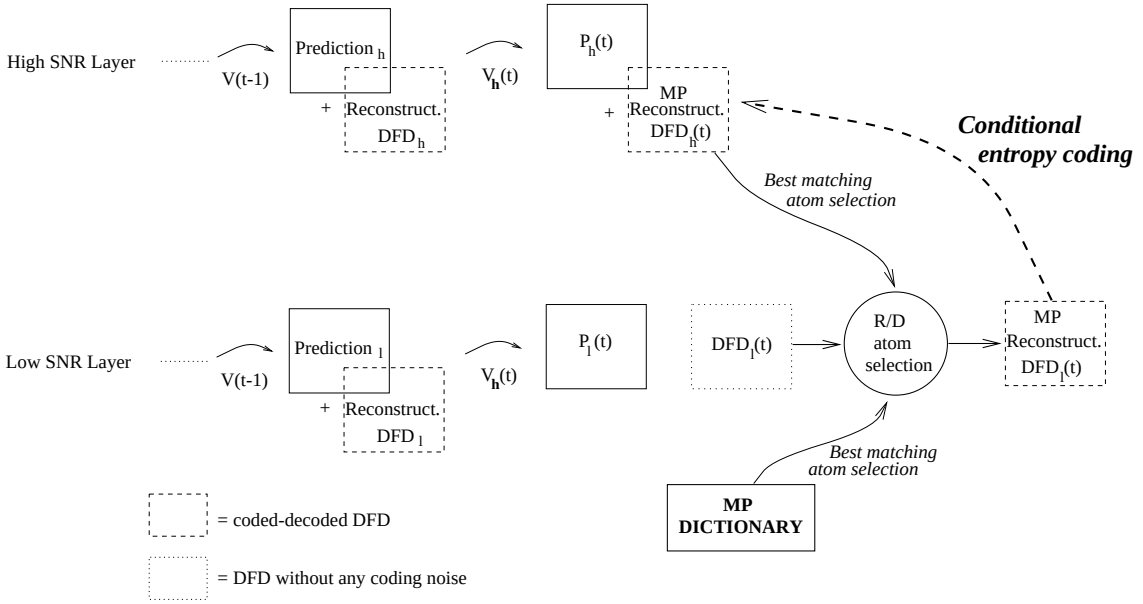

Figure 3: Construction of the low SNR layer DFD for optimal prediction of the high SNR layer DFD.

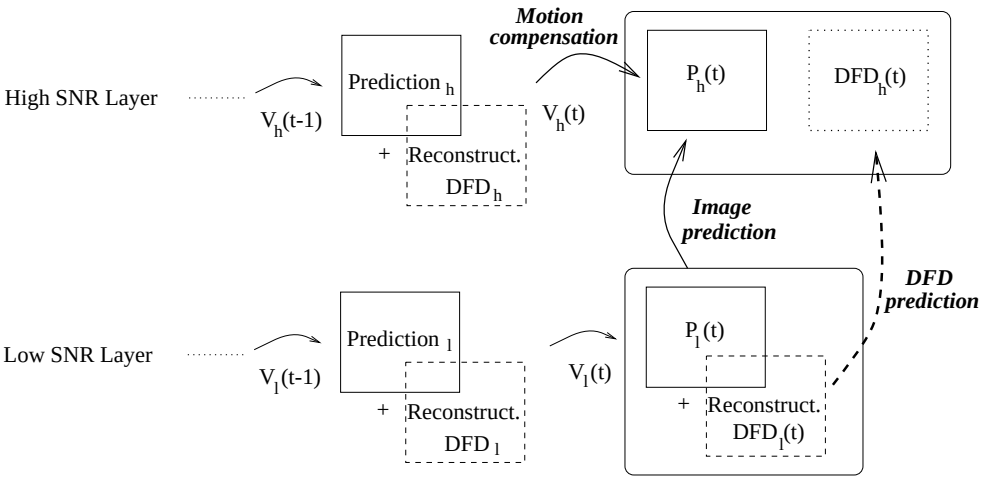

Figure 4: Multimodal MacroBlock-based prediction of the high SNR layer. 3 modes are considered: (a) MB is motion-compensated, (b) MB is predicted from the low SNR layer, or (c) MB is motion-compensated and the DFD is predicted from the low SNR layer reconstructed DFD.

Sequence	Layer	Simulcast with MP	Atom-based prediction	Block-based prediction	Frame prediction	MPEG-4
Hall	Low SNR	9053	9053	9053	9053	13594
	High SNR	16076	19510	20332	19605	33660
	Low+High	25129	19510	20332	19605	33660
Silent	Low SNR	22178	22178	22178	22178	25218
	High SNR	42419	52311	51397	51673	69667
	Low+High	64597	52311	51397	51673	69667
Foreman	Low SNR	47990	47990	47990	47990	49404
	High SNR	76515	94530	87460	86890	107424
	Low+High	124505	94530	87460	86890	107424
Coastguard	Low SNR	52804	52804	52804	52804	56676
	High SNR	88097	115697	112401	115837	141040
	Low+High	140901	115697	112401	115837	141040

Table 2: Low SNR layer, high SNR layer and both bit-rates layers (bits/s) for five schemes. The low SNR layer is encoded as in the non-scalable scheme. The set of motion vectors is estimated on the low SNR layer.

Sequence	Layer SNR	MacroBlock-based prediction schemes				
		Base layer motion vectors set			Two motion vectors sets	
		3 modes	MC, L-SNR-I	MC, MC+DFD	3 modes	MC, L-SNR-I
Hall	Low	9053	9053	9053	9053	9053
	High	20332	22586	20453	21058	21255
	Low + High	20332	22586	20453	21058	21255
Silent	Low	22178	22178	22178	22178	22178
	High	51397	53984	52384	50451	50739
	Low + High	51397	53984	52384	50451	50739
Foreman	Low	47990	47990	47990	47990	47990
	High	87460	92853	89004	88686	92253
	Low + High	87460	92853	89004	88686	92253
Coastguard	Low	52804	52804	52804	52804	52804
	High	112401	118123	117441	112686	118019
	Low + High	112401	118123	117441	112686	118019

Table 3: Comparison of the bit-rates (bits/s) required when considering 3 or 2 modes of prediction in a block-based prediction scheme. The low SNR layer is encoded as in the non-scalable scheme.

Sequence	Layer SNR	Simulcast with MP	Atom-based prediction	Frame-based prediction			MB-based prediction
				γ=0	γ=0.2	γ=1	
Hall	Low	9053	12132	9155	9642	14300	9108
	High	16073	17178	19477	18911	18341	19953
	Low + High	25126	17178	19477	18911	18341	19953
Silent	Low	22178	29331	22888	24186	37711	22584
	High	42418	44454	48376	47856	49914	48419
	Low + High	64596	44454	48376	47856	49914	48419
Foreman	Low	47990	57137	50012	53747	87848	49746
	High	76514	81652	82053	83809	104564	84196
	Low + High	124504	81652	82053	83809	104564	84196
Coastguard	Low	52804	65662	54621	56152	82129	54398
	High	88096	92956	108814	104031	92860	108622
	Low + High	140900	92956	108814	104031	92860	108622

Table 4: Low SNR layer, high SNR layer and both bit-rates layers (bits/s) for atom-based, MacroBlock-based and frame-based prediction schemes. The set of motion vectors is estimated on the high SNR layer.