Quality Scalable Video Coding

Vicente González Ruiz
Juan Pablo García Ortiz
Manuel Francisco López Martínez
Inmaculada García Fernández

Depto de Arquitectura de Computadores y Electrónica
Universidad de Almería

Barcelona, 27 de Febrero de 2009

Some initial notation

- Let V a video signal, a sequence of N digital images of $X \times Y$ pixels.
- We define that

$$V = \{V_i\}; 0 \leq i < N,$$

where V_i is the $i-$th image of V.

Why temporal scalability?

- Useful for a fast seek.
- $V^{(t)} = \{V_{2^t \times i}\}; 0 \leq i < \frac{N}{2^t}$ is the t-th temporal resolution level of V. Notice that $V = V^{(0)}$.

Why spatial scalability

- Ideal to show a high resolution video in low resolution displays.
- $V^{<s>}$ is the s-th spatial resolution level of V, i.e. the subsampled $\frac{Y}{2^s} \times \frac{X}{2^s}$ version of V $(0 \leq s < S)$. Notice that $V = V^{<0>}$.

Why quality scalability?

- Ideal for remote video browsing.
- $V^{[q]}$ is the q-th distortion level of V, i.e. the approximated reconstruction of V using $Q-q$ quality layers $(0 \leq q \leq Q)$. $V^{[0]} = V$ and $V^{[Q]} = 0$.

QSVC with Motion JPEG2000 [7]

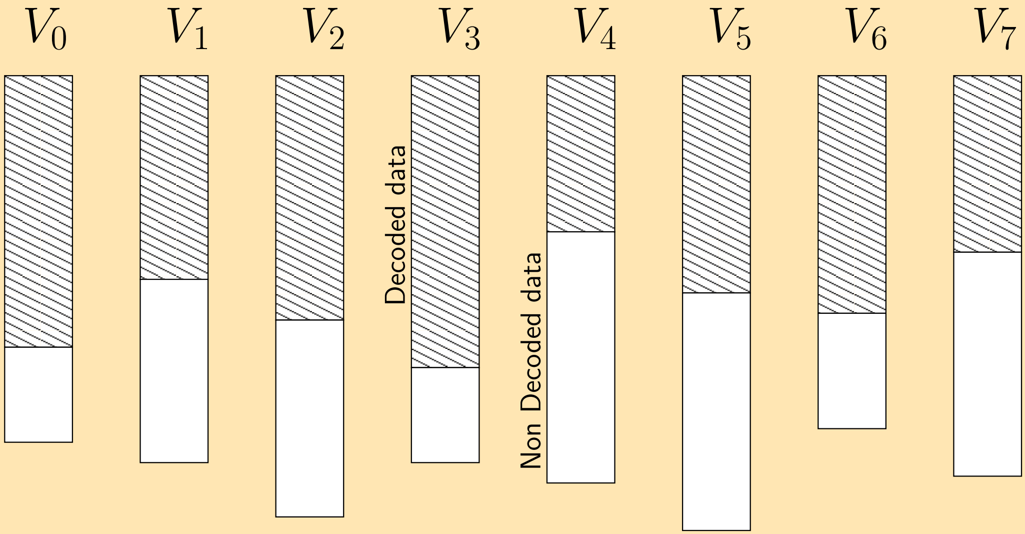
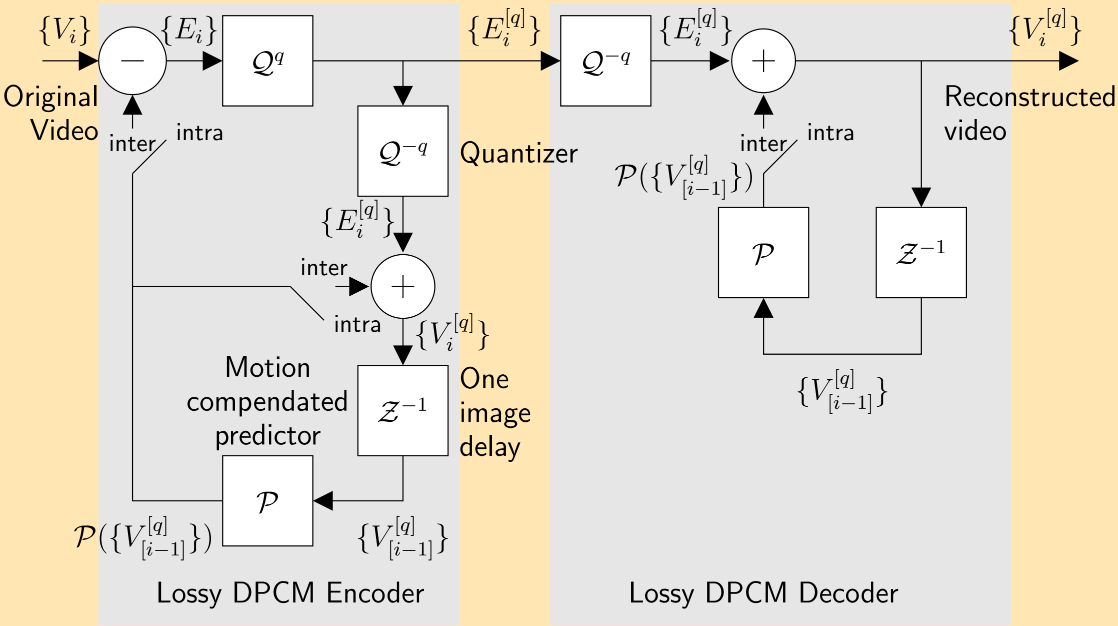


Image relationship: III$\cdots$.

Performance of Motion JPEG2000

The closed-loop approach [3, 1, 10]



The closed-loop approach

- From the previous figure we deduce that

$$E = \{V_0, H_1, H_2, \cdots\},$$

where $H_1, H_2, \cdots$ are the residual images.

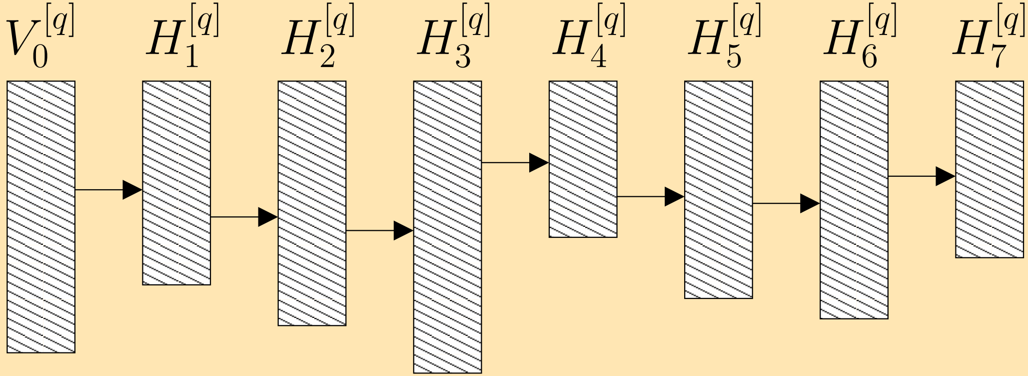
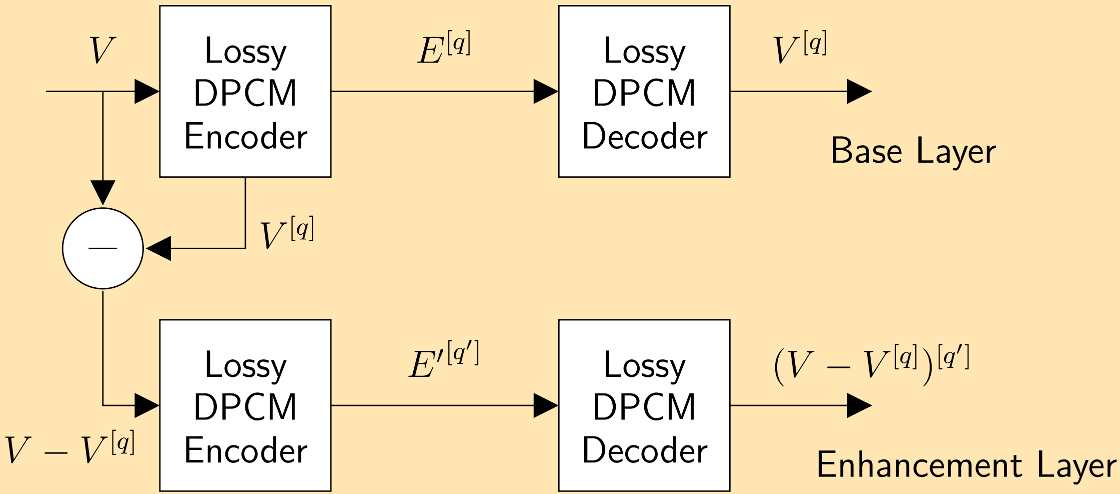


Image relationship: IPPP$\cdots$.

A 2-layers quality scalable codec



An example for 2 loops

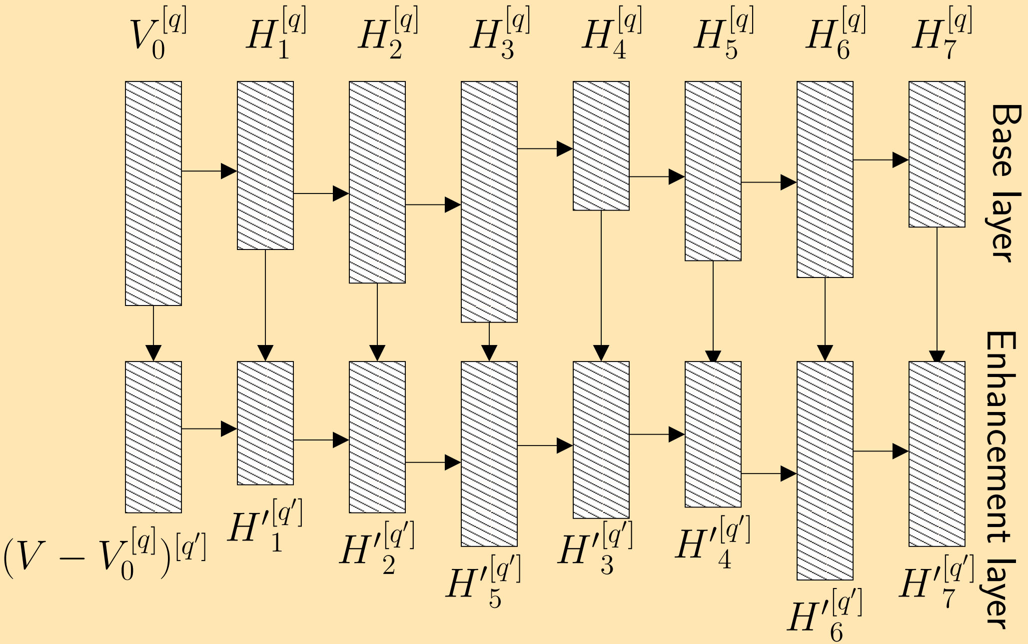
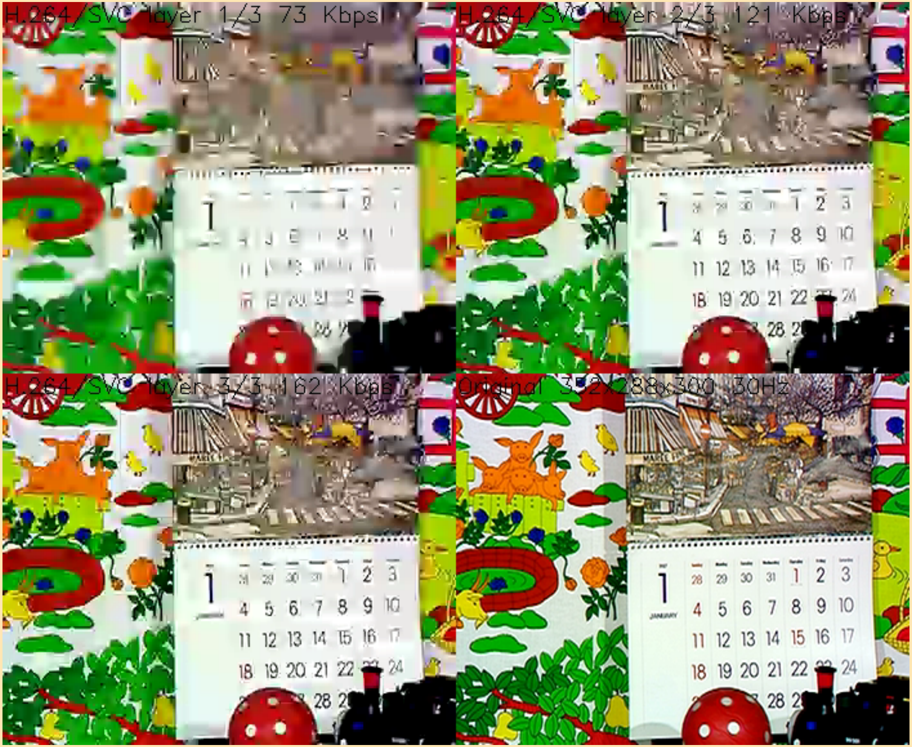
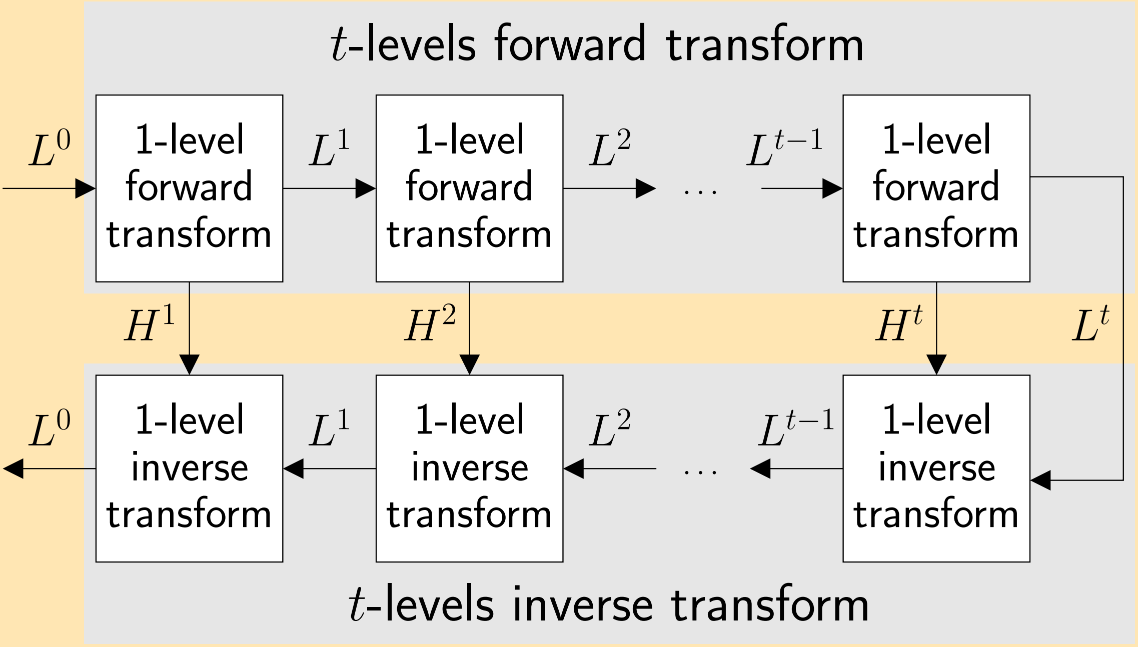


Image relationship: IPPP$\cdots$ in each loop (layer).

Performance of H.264/SVC (CGS) [4]



The open-loop (subband) approach [2]



The 1-level transform [6, 5, 9]

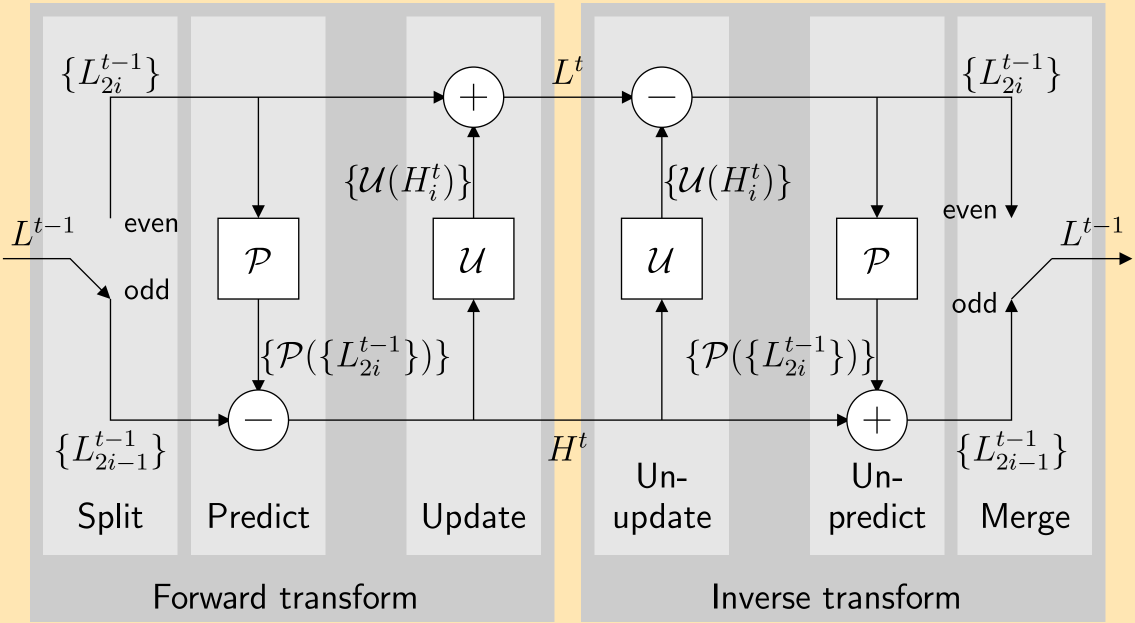


Image dependencies for SBC

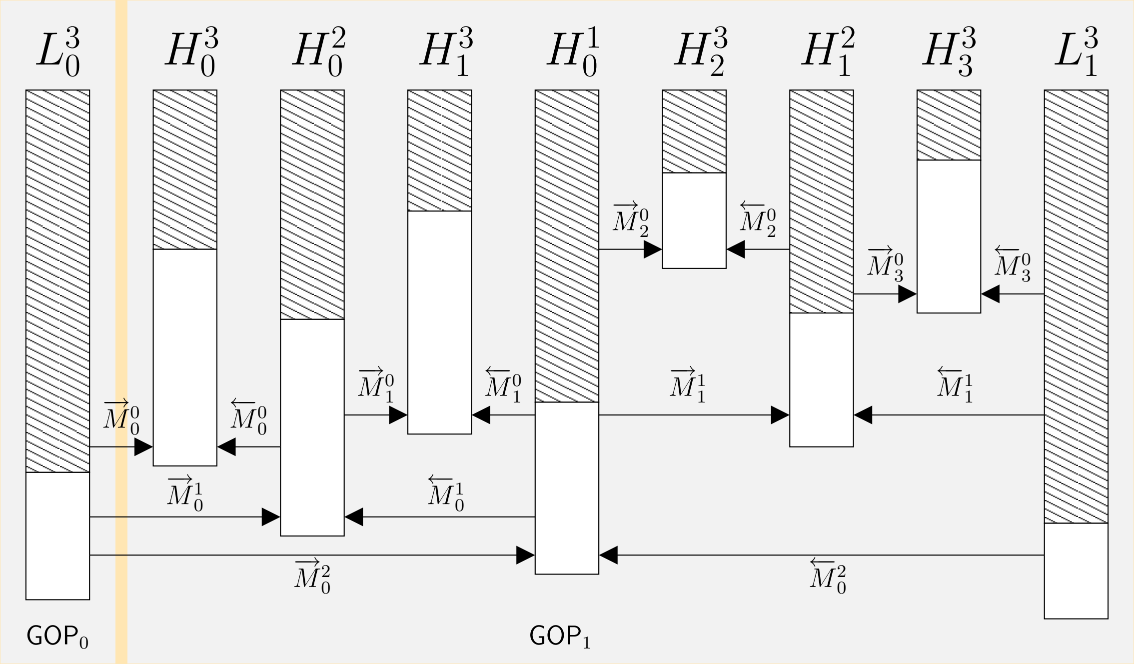


Image relationship: IBBBBBBBI.

The bit-rate allocation problem

- Let $b_j; j \in T = \{L^t, H^i; i = 1, \cdots, t\}$ the bit-rate of each subband, where

$$b = b_{L^t} + b_{H^t} + b_{H^{t-1}} + \cdots + b_{H^1}$$

  is the bit-rate of the encoded output texture data.

- Let

$$B = b + b_M$$

  the bit-rate of the encoded video, where b_M is the bit-rate of M, the motion data.

Bit-rate allocation for SNR scalability

- Any T data should be received after the M data.
- However, ¿how to sort (find the optimal ordering of the data of) the bit-streams of the T sub-bands in order to minimize the distortion of a progressive reconstruction of a GOP?

The $B = b = b_{L^t}$ case

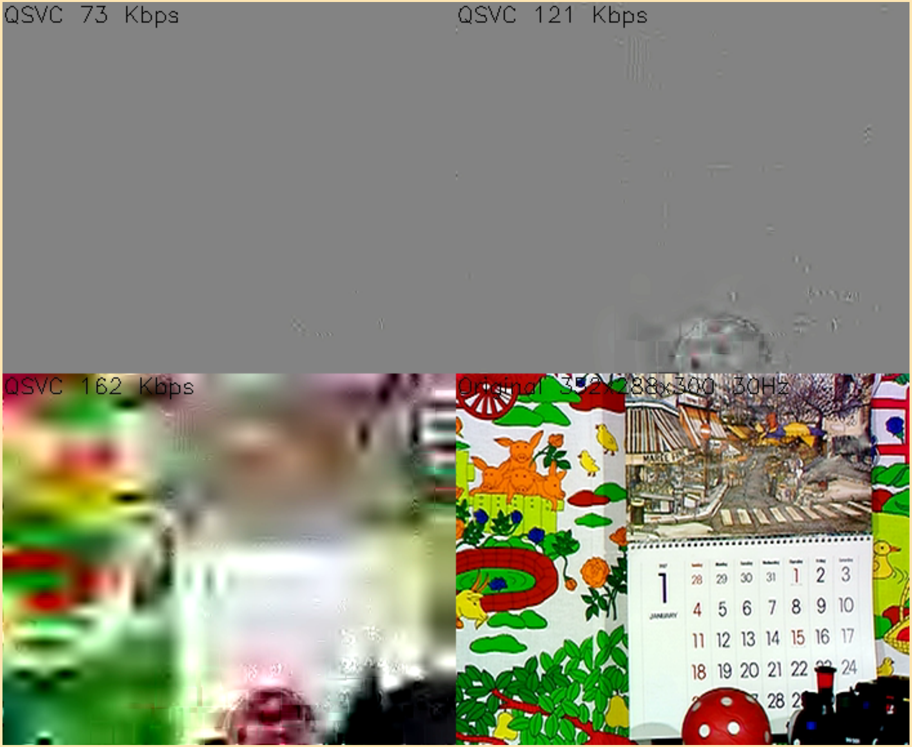
- A ideal bit-rate allocation scheme for a (rare) video with no movement ($M = 0$), where the predictions are identical to the predicted pictures and the residue subbands should be zero ($\{H^i = 0; i = 1, \cdots, t\}$).

The $B = b = b_{Lt}$ case

The $b_{L^t} = b_{H^t}; b_{H^{t-1}} = 2b_{H_t}$ case

- The opposite extreme case to the previous one is a video where each picture is completely different to the rest of the sequence. In this situation there is no correlation between pictures and the residues accumulate the information of the prediction and the predicted pictures. As the number of pictures in the residue subband H^{t-1} is twice the number of pictures in the subband H^t, we conclude that $b_{H^{t-1}} = 2b_{H_t}$.

The $b_{L^t} = b_{H^t}; b_{H^{t-1}} = 2b_{H_t}$ case



The $b_{L^t} = b_{H^t} = \cdots = b_{H^1}$ case

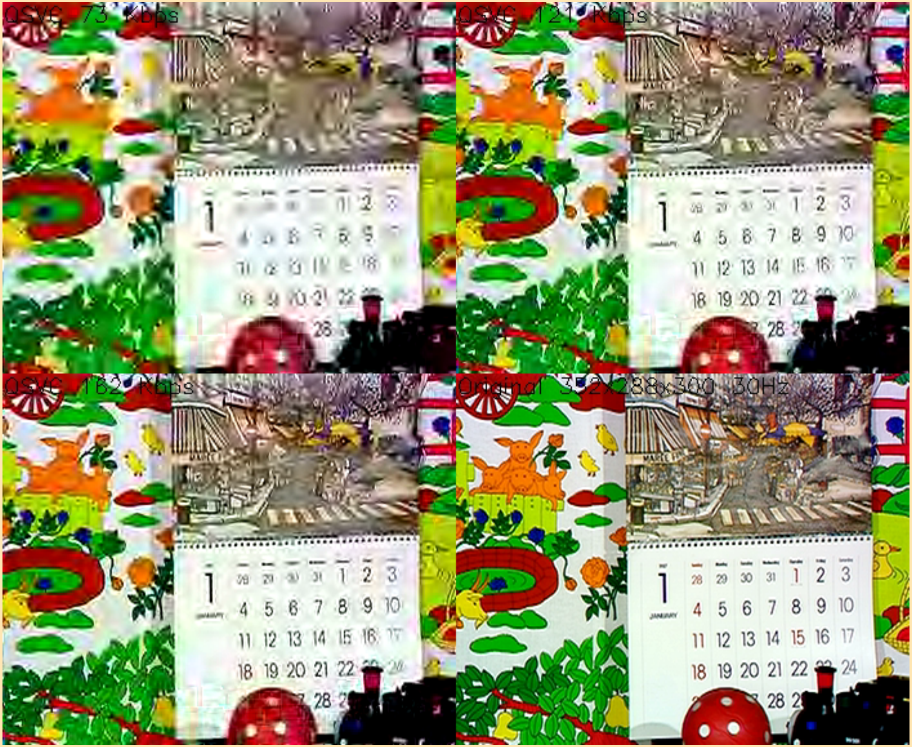
- In this bit-rate allocation strategy, the importance of a picture in the subbands L^t and H^t is the same, but the importance of a picture of the subband H^t is the double than the importance of a picture of the subband H^{t-1}.

The $b_{Lt} = b_{Ht} = \cdots = b_{H1}$ case

The $b_{L^t} = 2b_{H^t}; b_{H^t} = 2b_{H^{t-1}}$ case

- Finally, a scheme where the number of bits assigned to a t-th subband doubles the number of bits assigned to the $t-1$-th subband. This means that the weight of a picture in the t-th subband is four times the weight of a picture in the $t-1$-th subband.

The $b_{Lt} = 2b_{Ht}; b_{Ht} = 2b_{Ht-1}$ case



Conclusions about the bit-rate allocation

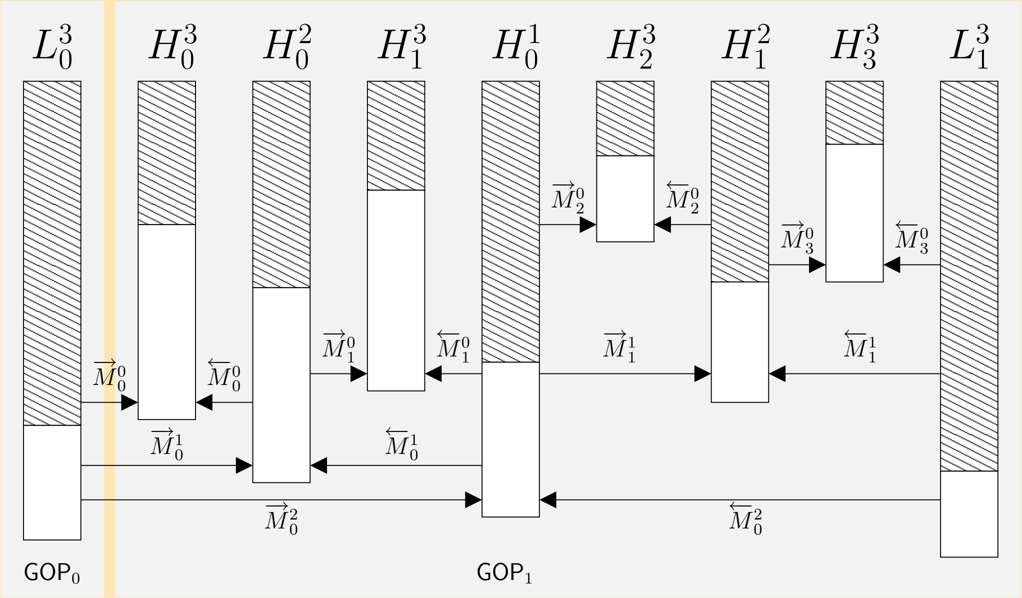
- The results show that the distortion of a encoded video depends on:
  - The characteristics of the video and the efficiency of the motion prediction model.
  - The relative weight (importance) of each temporal subband.
  - The relative weight of each image of each subband.
  - Moreover, the rate-distortion minimization performed by progressive codec (such as JPEG2000 [8]) for one image could not maximize the overall GOP's SNR.

Conclusions about the bit-rate allocation

- Static (predefined) bit-rate allocation strategies can not be optimal for any video.
- A special PCRD-opt for QSVC that takes in account all the information in T should be performed!

Motion data encoding

- M^t and M^{t-1} are correlated. $\overleftarrow{M}_i^t$ and $\overrightarrow{M}_i^t$ are correlated.



Motion data encoding

- In most sequences, it can be found that

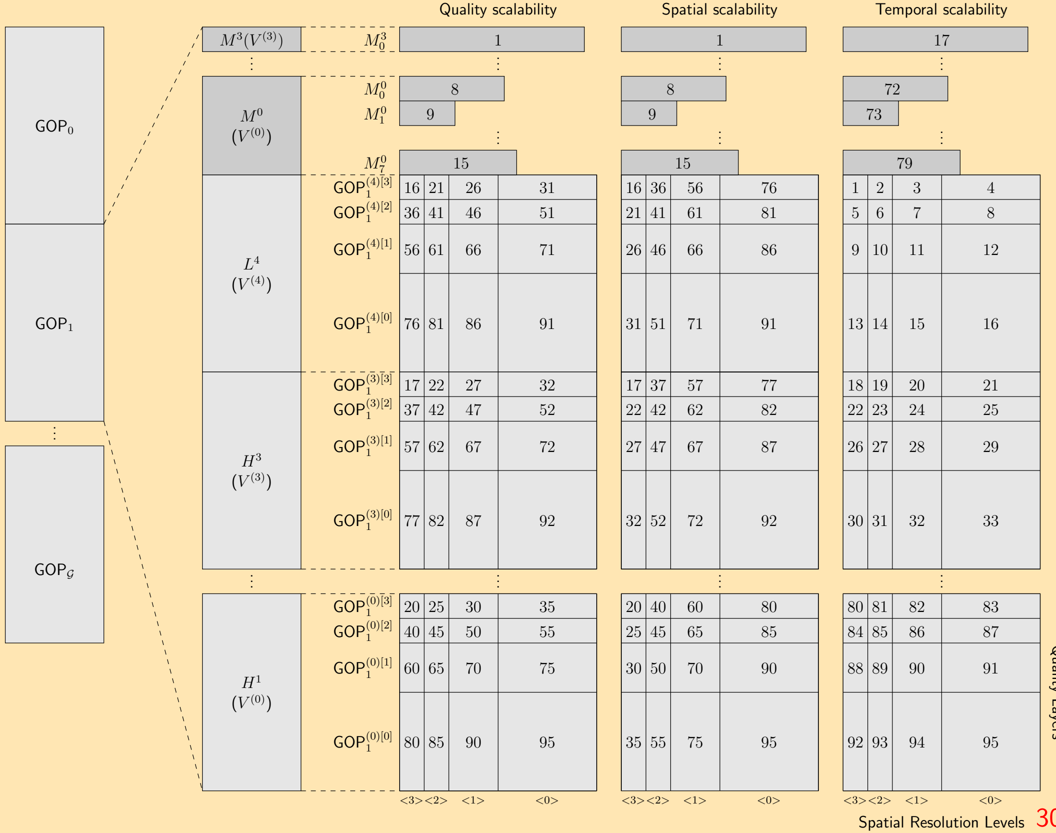
$$\overleftarrow{M}_i^t \approx -\overrightarrow{M}_i^t \tag{1}$$

and that

$$\overleftrightarrow{M}_i^t = 2\overleftrightarrow{M}_{2i}^{t-1}. \tag{2}$$

- Before the entropy encoding of M, Predictors 1 and 2 are applied.
- In the case that M^{t-1} is not received, it could be iferred from M^t that is transmitted first.
- Some degree of spatial redundancy can also be removed from M.

Stream organizations



Reconstruction of a truncated QSVC stream

- Encoding bit-rate: 1024 Kbps.
- Number of JPEG2000 layers: 16.
- Number of iterations of the temporal transform: 5.
- Number of iterations of the spatial transform: 5.
- Truncation pattern:

  GOP 1: 512 Kbps (64000 bytes).

  GOP 2: 64 Kbps (4000 bytes).

  GOP 3: 512 Kbps (66000 bytes).

  GOP 4: 64 Kbps (4000 bytes).

  ⋮

Reconstruction of a truncated QSVC stream

References

[1] Bernd Girod. The efficiency of motion-compensating prediction for hybrid coding of video sequences. *IEEE Journal on Selected Areas in Communications*, SAC-5(7):1140–1154, 1987.

[2] J.-R. Ohm. Three-dimensional subband coding with motion compensation. *IEEE Transactions on Image Processing*, 3:559–571, 1994.

[3] Karel Rijkse. H.263: Video coding for low-bit-rate communication. *IEEE Communications Magazine*, 34(12):42–45, 1996.

[4] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. Overview of the scalable video coding extension of the h.264/avc standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, September 2007.

[5] Andrew Secker and David Taubman. Lifting-based invertible motion adaptive transform (limat) framework for highly scalable video compression. *IEEE Transactions on Image Processing*, 12(12):1530–1542, December 2003.

[6] Wim Sweldens. The Lifing Scheme: A Custom-Desing Constuction of Biorthogonal Wavelets. *Applied and Computational Harmonic Analysis*, 3(2):186–200, 1996. http://cm.bell-labs.com/who/wim/papers/lift1.pdf.

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

[8] David Taubman. High performance scalable image compression with ebcot. *IEEE Transactions on Image Processing*, 9(7):1158–1170, 2000.

[9] David Taubman, Reji Mathew, and Nagita Mehrseresht. Fully Scalable Video Compression with Sample-Adaptive Lifting and Overlaped Block Motion. In A. Said and J. G. Apostolopoulos, editors, *Proceeding of SPIE-IS&T Electronic Imaging*, volume 5685, pages 366–377, March 2005.

[10] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h.264/avc video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576, 2003.