AN IMPROVED PYRAMID FOR SPATIALLY SCALABLE VIDEO CODING

Markus Flierl and Pierre Vandergheynst

Signal Processing Institute
Swiss Federal Institute of Technology Lausanne
{markus.flierl,pierre.vandergheynst}@epfl.ch

ABSTRACT

This paper discusses an improved pyramid for spatially scalable video coding. We introduce additional update steps in the analysis and the synthesis of the Laplacian pyramid. Our pyramid is able to control efficiently the quantization noise energy in the reconstruction. Hence, it provides improved coding performance when compared to the standard Laplacian pyramid. Moreover, our pyramid does not require biorthogonal filters as they should be used for the frame reconstruction of the Laplacian pyramid. Therefore, low-pass filters can be chosen that suppress aliasing in the low-resolution images efficiently and, hence, permit efficient motion compensation. The experimental results demonstrate coding gains of up to 1 dB for both images and image sequences when compared to the standard Laplacian pyramid.

1. INTRODUCTION

Spatial scalability of video signals can be achieved with critically sampled spatial wavelet schemes but also with an overcomplete spatial representation. Critically sampled schemes struggle with the problem that critically sampled high-bands are shift-variant. Therefore, efficient motion compensation is challenging. On the other hand, overcomplete representations can be shift-invariant, thus permitting efficient motion compensation in the spatial subbands, but they have to be designed carefully to achieve high compression efficiency.

This paper aims to improve the coding efficiency of overcomplete spatial representations for video coding. The Laplacian pyramid proposed by Burt and Adelson [1] provides such an overcomplete multiresolution representation. In [2], the Laplacian pyramid is treated as a frame operator. When using the dual frame operator for the reconstruction, its compression efficiency can be improved. But these framed pyramids require biorthogonal filters if the reconstruction shall be an inverse of the Laplacian pyramid.

Biorthogonal filters in the framed pyramid may cause significant aliasing in the low-resolution pictures. These aliasing components burden efficient motion compensation in the spatial low-bands and may degrade overall video coding performance.

This paper proposes a so called "lifted pyramid" that improves the Laplacian pyramid scheme but does not require biorthogonal filters like the framed pyramid. In particular, the resulting spatial subbands can be efficiently coded with motion-compensated temporal transforms [3, 4, 5]. The combination of a lifted pyramid with motion-compensated temporal transforms on the spatial subbands provides rate-distortion efficient spatial and temporal scalability for video signals.

The outline of the paper is as follows: Section 2 revises the Laplacian pyramid and provides the basis for the lifted pyramid in Section 3. Section 4 motivates the new scheme by discussing the reconstruction with ideal low-pass filters. Experimental results for images and image sequences are presented in Section 5.

2. LAPLACIAN PYRAMID

The Laplacian pyramid (**Fig. 1**) provides a method for multiresolution data representation [1]. The basic idea is as follows: First, a coarse approximation of the original signal is low-pass filtered and downsampled. The coarse version is then used to provide a prediction signal by upsampling and filtering to calculate the prediction error with respect to the original. For the synthesis, the reconstructed signal is obtained by simply adding back the prediction error to the prediction from the coarse signal.

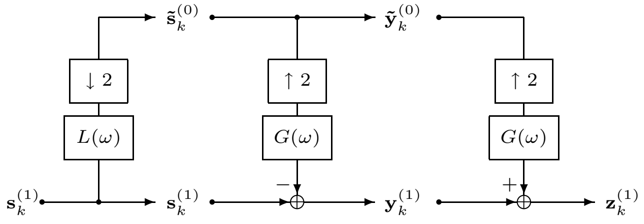

Fig. 1. Pyramid scheme. The high-resolution image $\mathbf{s}_k^{(1)}$ is filtered by the filter $L(\omega)$ and downsampled by factor 2 to generate the low-resolution image $\tilde{\mathbf{s}}_k^{(0)}$. Analysis and synthesis use this low-resolution image $\tilde{\mathbf{y}}_k^{(0)}$ to form a high-resolution prediction image. The analysis subtracts this prediction image and outputs the high-resolution difference image $\mathbf{y}_k^{(1)}$.

The multiresolution representation of the Laplacian pyramid is overcomplete. That is, there are more coefficients after the analysis than in the input. In particular, this is a burden for coding applications, where additional quantization noise energy degrades the reconstruction. To control the quantization noise energy in the reconstruction, we extend analysis and synthesis by an update of the coarse signal.

3. LIFTED PYRAMID

In contrast to the Laplacian pyramid, we update the coarse signal at analysis and synthesis by filtering and downsampling the detail signal. The prediction step of the Laplacian pyramid with the new update step forms a sequence of spatial lifting steps. This "lifted" pyramid scheme is shown in **Fig. 2**. The high-resolution image

$\mathbf{s}_k^{(1)}$ is filtered by $L(\omega)$ and downsampled by factor 2 to generate the low-resolution image $\tilde{\mathbf{s}}_k^{(0)}$. Analysis and synthesis upsample this low-resolution image $\tilde{\mathbf{y}}_k^{(0)}$ by 2 and filter with $G(\omega)$ to form a high-resolution prediction image. The analysis subtracts this prediction image and outputs the high-resolution difference image $\mathbf{y}_k^{(1)}$. The update step filters the high-resolution difference image $\mathbf{y}_k^{(1)}$ with $U(\omega)$ and downsamples by factor 2. At the analysis, this update signal is used to generate a low-resolution low-band $\tilde{\mathbf{y}}_k^{(0)}$. Note, the scheme is reversible and permits perfect reconstruction for any set of filters used.

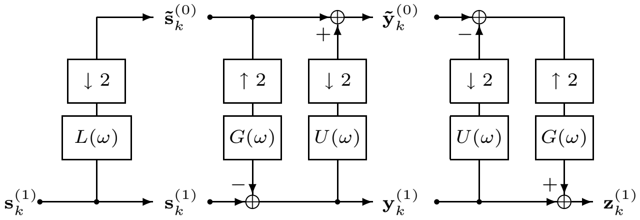

Fig. 2. Lifted pyramid scheme. The pyramid scheme extends analysis and synthesis by a spatial update step which filters the high-resolution difference image $\mathbf{y}_k^{(1)}$ with $U(\omega)$ and downsamples by factor 2. At the analysis, this update signal is used to generate a low-resolution low-band $\tilde{\mathbf{y}}_k^{(0)}$.

The lifted pyramid is reversible for any set of filters $L(\omega)$, $G(\omega)$, and $U(\omega)$ due to the lifting structure. An interesting special case is given if $L(\omega) = U(\omega) := H(\omega)$ and if $H(\omega)$ and $G(\omega)$ are biorthogonal with respect to the sampling lattice 2. In that case, the resulting update signal at the analysis is zero and we obtain the framed pyramid of [2] as depicted in **Fig. 3**. [2] shows that the synthesis in **Fig. 3** is an inverse transform of the Laplacian pyramid if and only if the two filters $H(\omega)$ and $G(\omega)$ are biorthogonal with respect to the sampling lattice 2.

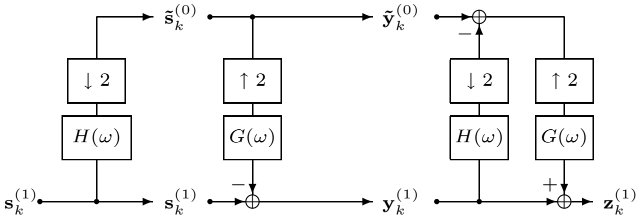

Fig. 3. Framed pyramid according to [2]. The spatial update step is used at the synthesis only. Therefore, the reconstruction is an inverse transform of the analysis if and only if the two filters $H(\omega)$ and $G(\omega)$ are biorthogonal with respect to the sampling lattice 2.

Using a biorthogonal filter for $L(\omega)$ may not be advisable for spatially scalable video coding. For example, the 9/7 biorthogonal filter $H(\omega)$ causes significant aliasing in the downsampled low-band. **Fig. 4** compares the frequency response of the low-pass filter $L(\omega)$ with its coefficients in **Table 1** to that of the 9/7 biorthogonal filter $H(\omega)$.

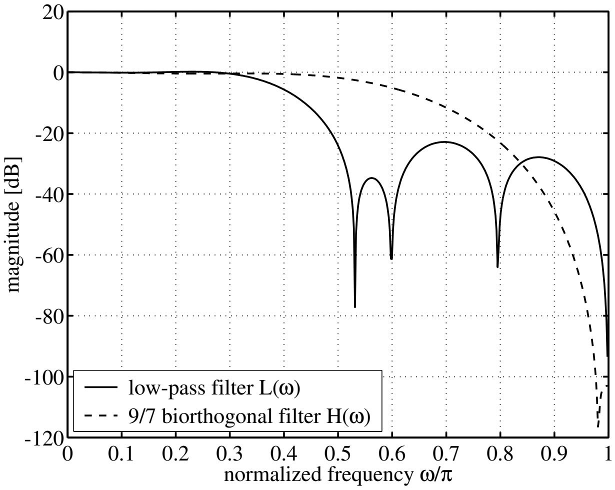

Fig. 4. Frequency responses of the low-pass filter $L(\omega)$ and the 9/7 biorthogonal filter $H(\omega)$. The 9/7 biorthogonal filter causes significant aliasing in the downsampled low-band $\tilde{\mathbf{y}}_k^{(0)}$.

n	0	±1	±2	±3	±4	±5	±6
$l[n]$	26/64	19/64	5/64	-3/64	-4/64	0	2/64

Table 1. Coefficients for the separable low-pass filter $L(\omega)$.

Fig. 5 depicts a detail of the image *Barbara* and visualizes this aliasing. For spatially scalable video coding, the aliasing in spatial low-bands burdens motion-compensated coding of the low-resolution images. Therefore, low-pass filters should be used that suppress aliasing components efficiently. Controlling both aliasing components in the low-resolution images and quantization noise energy in the reconstruction, we propose the lifted pyramid for spatially scalable video coding.

Fig. 5. Detail of image *Barbara*. The original is filtered and downsampled. The low-pass filter $L(\omega)$ removes high-frequency details (left) whereas the 9/7 filter $H(\omega)$ causes aliasing (right).

4. RECONSTRUCTION WITH IDEAL LOW-PASS

In the following, we discuss briefly the propagation of the quantization noise energy in the synthesis and the impact on the reconstructed image. For that, we choose for the update and the prediction filters the ideal low-pass filter $U(\omega) = G(\omega) = \mathbf{1}_{\mathcal{B}}(\omega)$,

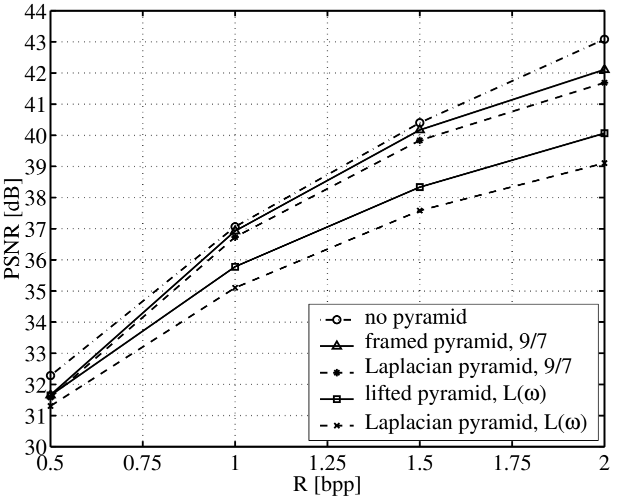

Fig. 6. Rate-distortion performance of the 1-level lifted pyramid for the image *Barbara* at resolution 512×512. The performance without any pyramid as well as with the Laplacian pyramid is given for reference. 9/7 wavelet and low-pass $L(\omega)$ are used.

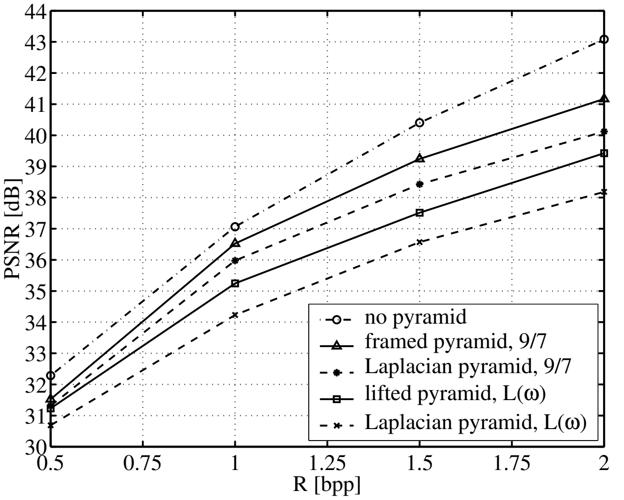

Fig. 7. Rate-distortion performance of the 2-level lifted pyramid for the image *Barbara* at resolution 512×512. The performance without any pyramid as well as with the Laplacian pyramid is given for reference. 9/7 wavelet and low-pass $L(\omega)$ are used.

which is one in the base-band $\mathcal{B} = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ and zero elsewhere. Further, we describe the low-resolution images as band-limited signals in the base-band $\mathcal{B}$.

According to **Fig. 1**, the spectrum of the reconstructed image for the Laplacian pyramid is

$$\mathbf{Z}(\omega) = G(\omega)\mathbf{Y}^{(0)}(\omega) + \mathbf{Y}^{(1)}(\omega). \qquad (1)$$

If the power spectral density (PSD) of the quantization noise $\Phi_{\mathbf{nn}}(\omega)$ is white with the same variance for low- and high-band, the PSD of the reconstructed image is $\Phi_{\mathbf{zz}}(\omega) = \Phi_{\mathbf{ss}}(\omega) + [1 + \mathbf{1}_{\mathcal{B}}(\omega)]\Phi_{\mathbf{nn}}(\omega)$.

For the lifted pyramid in **Fig. 2**, the spectrum of the reconstructed image is

$$\mathbf{Z} = G(\omega)\mathbf{Y}^{(0)}(\omega) + [1 - G(\omega)U(\omega)]\,\mathbf{Y}^{(1)}(\omega). \qquad (2)$$

With the same assumptions for the quantization noise, the PSD of the reconstructed image is $\Phi_{\mathbf{zz}}(\omega) = \Phi_{\mathbf{ss}}(\omega) + \Phi_{\mathbf{nn}}(\omega)$. That is, the lifted pyramid is able to suppress the quantization noise in the base-band. This fact improves the coding efficiency of the lifted pyramid.

5. EXPERIMENTS

We investigate the coding efficiency of the lifted and framed pyramid with 1 and 2 decomposition levels. In addition, we compare to the Laplacian pyramid as well as to coding without any pyramid scheme. We decompose the image *Barbara* at resolution 512×512 and the first picture of the image sequence *City* at 4CIF resolution. The resulting subbands are encoded with the JPEG 2000 image coding standard [6].

Figs. 6 and **7** show the coding efficiency of the 1- and 2-level pyramids for the image *Barbara*, respectively. If no pyramid is used, the image is coded directly with JPEG 2000. The framed pyramid in **Fig. 3** uses the 9/7 biorthogonal filters. In addition, the Laplacian pyramid in **Fig. 1** is also given when using the 9/7 biorthogonal filters. Finally, the lifted pyramid uses for all down- and upsampling filters the low-pass $L(\omega)$ with the coefficients in **Table 1**. Again, the Laplacian pyramid in **Fig. 1** is also given when using the low-pass $L(\omega)$.

Similar to the image *Barbara*, **Figs. 8** and **10** depict the coding efficiency of the 1- and 2-level pyramids for the first picture of the image sequence *City*, respectively. We observe for both the lifted pyramid with the low-pass filter $L(\omega)$ and the framed pyramid with the 9/7 biorthogonal filter that the additional update step improves the coding efficiency. When compared to the 1-level pyramids, the relative improvements are slightly larger for the 2-level pyramids which have more subband samples and, hence, more quantization noise energy. Due to the biorthogonality of the 9/7 filters, the coding efficiency is more advantageous when compared to that of the low-pass $L(\omega)$. But the pyramid with the 9/7 biorthogonal filters is burdened by significant aliasing components in the low-resolution images. This is not the case when using the low-pass $L(\omega)$. Moreover, the coding efficiency can be improved by using more accurate low-pass filters.

Finally, we present the video coding efficiency of the lifted pyramid with 1 decomposition level and compare to the Laplacian pyramid. We decompose 120 pictures of the image sequence *Container Ship* at 30 fps in CIF resolution with the lifted pyramid. The resulting QCIF and CIF image sequences are encoded with the MCTF part of the Joint Scalable Video Model (JSVM) [7]. That is, the spatial scalability provided by the JSVM is not used.

Fig. 9 shows the coding efficiency of the reconstructed sequence *Container Ship* at 30 fps in CIF resolution when coded with the lifted pyramid and the Laplacian pyramid. Both pyramids use the low-pass $L(\omega)$ to suppress efficiently aliasing components in the low-resolution images. The lifted pyramid provides gains of almost 1 dB over the Laplacian pyramid. The rate-distortion performance of the sub-streams that represent the sequence *Container Ship* at 30 fps in QCIF resolution is also shown. Note that the low-resolution representation of the lifted pyramid requires a larger bit rate when compared to that of the Laplacian pyramid at

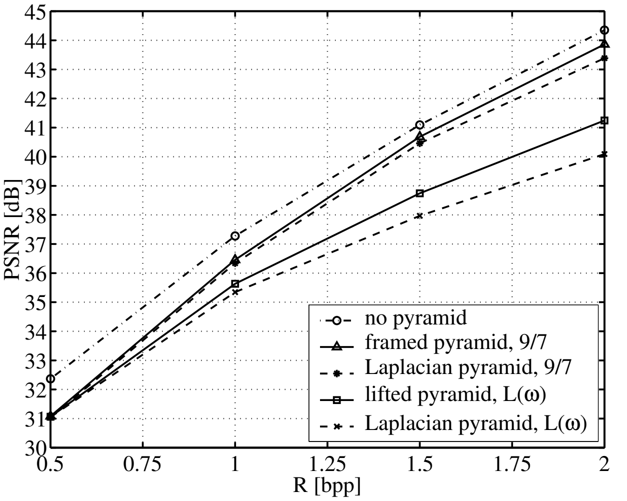

Fig. 8. Rate-distortion performance of the 1-level lifted pyramid for the image *City* at 4CIF resolution. The performance without any pyramid as well as with the Laplacian pyramid is given for reference. 9/7 wavelet and low-pass $L(\omega)$ are used.

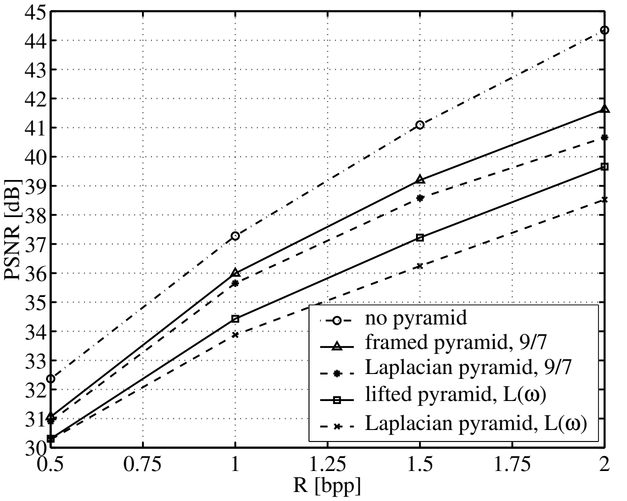

Fig. 10. Rate-distortion performance of the 2-level pyramid for the image *City* at 4CIF resolution. The performance without any pyramid as well as with the Laplacian pyramid is given for reference. 9/7 wavelet and low-pass $L(\omega)$ are used.

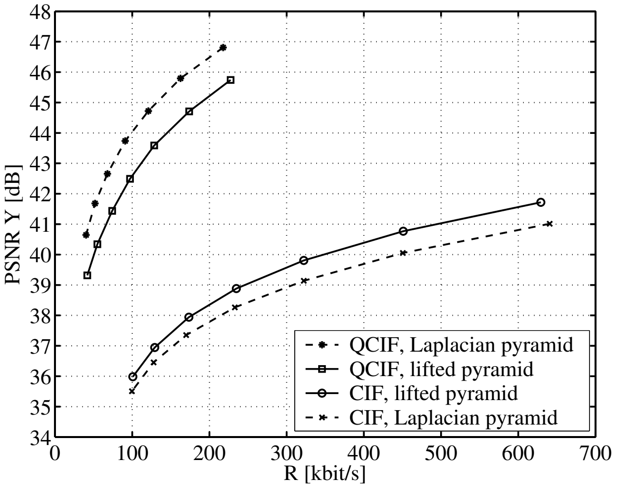

Fig. 9. Rate-distortion performance of the reconstructed sequence *Container Ship* at 30 fps in CIF resolution. The lifted pyramid is compared to the Laplacian pyramid. The performance of the sub-streams that represent the sequence in QCIF resolution is also shown. $\text{QP}^{(1)} = \text{QP}^{(0)} + 6$ holds for the quantization parameters.

the same quality. Due to the spatial update step, the low-resolution representation has more significant high-frequency components.

6. CONCLUSIONS

We discussed an improved pyramid for spatially scalable video coding. The additional update step in the analysis and the synthesis results in a lifted pyramid. This pyramid is able to control efficiently the quantization noise energy in the reconstruction. Hence, it provides improved coding performance when compared to the standard Laplacian pyramid. Moreover, the lifted pyramid does not require biorthogonal filters as the framed pyramid. With the lifted pyramid, low-pass filters can be chosen that suppress aliasing efficiently and, hence, permit efficient motion compensation.

7. ACKNOWLEDGMENT

The authors would like to thank Jose Luis Fernández for helpful discussions and for providing the coding results with the JSVM.

8. REFERENCES

[1] P.J. Burt and E.H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[2] M.N. Do and M. Vetterli, “Framing pyramids,” *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2329–2342, Sept. 2003.

[3] B. Pesquet-Popescu and V. Bottreau, “Three-dimensional lifting schemes for motion compensated video compression,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City, UT, May 2001, vol. 3, pp. 1793–1796.

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

[5] M. Flierl and B. Girod, “Video coding with motion-compensated lifted wavelet transforms,” *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 561–575, Aug. 2004.

[6] *An Implementation of the JPEG 2000 Standard in Java*, http:// jj2000. epfl. ch, Oct. 2002, version 4.1.

[7] Joint Video Team, *Joint Scalable Video Model JSVM 1*, Jan. 2005, http:// ftp3. itu. int/ av-arch/ jvt-site/ 2005_01_Hong Kong/ JVT-N023.zip.