Wavelet-Based Fast Decoding of 360° Videos
– Supplementary Material –

Colin Groth (iD), Sascha Fricke (iD), Susana Castillo (iD), and Marcus Magnor (iD)

✦

1 FUNDAMENTALS

1.1 Introduction to the wavelet theory

The idea to transform signals with wavelets was founded in the 1980s. With a better balance between temporal and frequency resolution than the Fourier transform, the wavelet theory showed a lot of potential and soon was further developed. This progression led to the development of the JPEG2000 standard, which is based on the discrete wavelet transformation. Although wavelet-based codecs for images and videos have never been adopted by the general public, wavelets are nowadays used by default for other purposes, e.g. for storing fingerprints by the FBI [4].

Grossmann and Morlet were the first to raise the concept of wavelet-based data analysis in their work from 1984 [1]. While working on the analysis of transient seismic signals, Morlet replaced the window function of a Gabor analysis (windowed fourier transform) with a function ψ, which is well defined in its temporal and spatial resolution. Instead of shifting the window in the frequency domain, as in the windowed fourier transform, the wavelet transform scales the window. In their work, the wavelet transform was shown to be a suitable method for the analysis of seismic signals with a better performance than the fourier transform. In the course of their studies, the name 'wavelet' arose, which describes the analysis function ψ. Despite this success, it soon became apparent that this continuous wavelet transform cannot be applied in a standardised form and is not always invertible. In 1985 Y. Meyer found out that for certain discrete values for a and b (scale and shift) an orthonormal basis (Hilbert basis) can be created for the Hilbert space $L^2(\mathbb{R})$ of the square integrable functions. Here the wavelet basis (also called mother-wavelet) is given by the function:

$$\psi \in L^2(\mathbb{R})$$

Among the most popular and widespread wavelets today are the Daubechies wavelets. The Daubechies wavelet family was constructed in 1988 by Ingrid Daubechies, who combined the common wavelet theory with the filter theory. Daubechies wavelets have compact support and are differentiable up to a prescribed finite order. Wavelets with compact support that are infinitely differentiable do not exist.

1.2 The Fast Wavelet Transform

When images are transformed with a wavelet transform, typically a fast wavelet transform (FWT) is used. The FWT is a specific use case of the wavelet transform that uses filter banks. For the FWT the image is convoluted with two wavelet kernels, a low and a high pass filter. These 1D kernels define the wavelet and represent the filter coefficients. With the high pass filter the frequency information (wavelet coefficients) is extracted. The low pass filter breaks the abstraction down to a lower resolution level. This process is recursively repeated on the

- All authors are with the Institute for Computer Graphics at TU Braunschweig, Germany.
  E-mail: {groth, fricke, castillo, magnor}@cg.cs.tu-bs.de.

approximation image until the desired level is reached. In theory this process can be repeated until only one pixel is left for the approximation image. To preserve the amount of information, the approximation and wavelet coefficients are kept only for every second pixel (in the transform dimension). Nevertheless, the information of the other pixels are preserved through the spanned filter banks. Therefore, the image can be reconstructed in the original quality.

In our case the 2D transform is first performed in x dimension. After transforming the whole image in the X dimension, the wavelet transform is performed on the already transformed image in the other dimension (y dimension). This order is reversible.

1.3 Inverse Fast Wavelet Transform

The original frames are reconstructed from the wavelet-based representation by an inverse version of the previously described wavelet transform. For the transform image, we saved a low-dimensional approximation of the image together with the coefficients of each wavelet. At this point, we leave out any compression for simplicity. As the name suggests, the inverse wavelet transform processes the frequencies in the opposite direction than the forward wavelet transform, starting with the lowest frequency and ending with the highest frequency. Also the order in which the two axes of the image are inverse transformed per frequency level has to be inverted. The inverse wavelet transform uses its own filter kernels for the low-pass and high-pass filtering. These kernels, or wavelet filters, for most wavelets look almost similar to their forward transform counterparts but with a reversed order of the values. All wavelets used in our work are widely known and therefore their kernels are clearly defined.

For the reconstruction of a specific pixel for some level, the low-pass filter convolutes the image approximation and simultaneously the high-pass filter is applied on the associated level frequency information which is represented by the wavelet coefficients. Both of the filters are centred over the pixel to be reconstructed. After the filter convolution, the two convolution values are summed up to form the approximation value of the next higher level. This process is performed recursively over all levels. After the transform of every level, the image approximation doubles in pixel size. After the transform of the last level the approximation of the image is the full reconstruction of the original image in its former resolution.

2 SCENES

For our experiment we used computer-generated videos as well as 360° recordings with static and moving camera trajectory. In particular eight videos with diverse scenes are used. The CGI scenes are rendered out of Unreal Engine 5 as 360° videos in native 8k stereo resolution. They contain:

1. an urban scene in a virtual New York City

2. a nature scene in a mountain area

The real-world moving camera scenes are originally for our previous work [2, 3] and cover:

1. downhill ride with a mountain bike

2. horse riding

3. wall climbing at an artificial wall

4. walking through a park in the summer

Furthermore, we use two static scenes with different lighting from related work [5]:

1. recording in a cave with little light

2. recording on a boat in the waters of Norway at daytime

In the Figures 1 – 8 we provide representative frames of all scene used for the experiment.

REFERENCES

[1] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736, 1984. doi: 10.1137/0515056 1

[2] C. Groth, J.-P. Tauscher, N. Heesen, S. Grogorick, S. Castillo, and M. Magnor. Mitigation of cybersickness in immersive 360°videos. In *IEEE Virtual Reality Workshop on Immersive Sickness Prevention (WISP)*, pp. 169–177. IEEE, 2021. doi: 10.1109/VRW52623.2021.0 1

[3] C. Groth, J.-P. Tauscher, N. Heesen, M. Hattenbach, S. Castillo, and M. Magnor. Omnidirectional galvanic vestibular stimulation in virtual reality. *Transactions on Visualization and Computer Graphics (TVCG)*, 28(5):2234–2244, 2022. doi: 10.1109/TVCG.2022.315 1

[4] B. Hubbard. *The World According to Wavelets: The Story of a Mathematical Technique in the Making, Second Edition*. CRC Press, 1998. 1

[5] M. Mühlhausen, M. Kappel, M. Kassubeck, P. M. Bittner, S. Castillo, and M. Magnor. Temporal consistent motion parallax for omnidirectional stereo panorama video. In *ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2020. doi: 10.1145/3385956.3418965 2
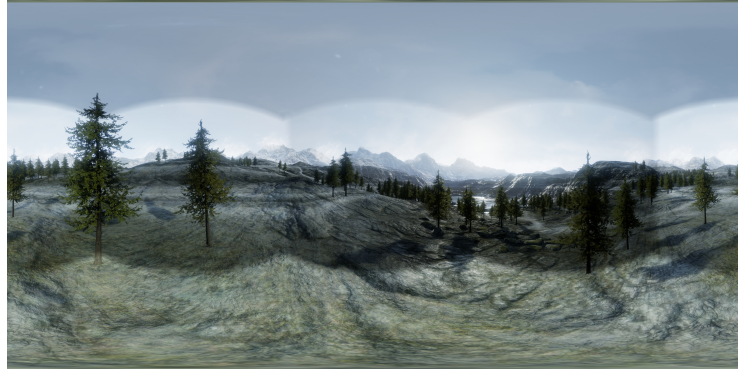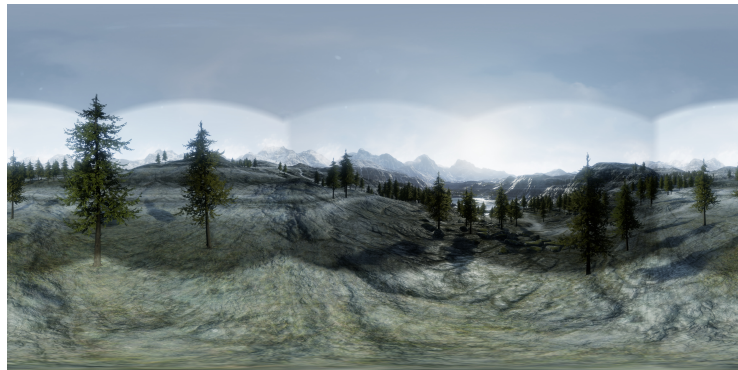
Fig. 1: The city scene
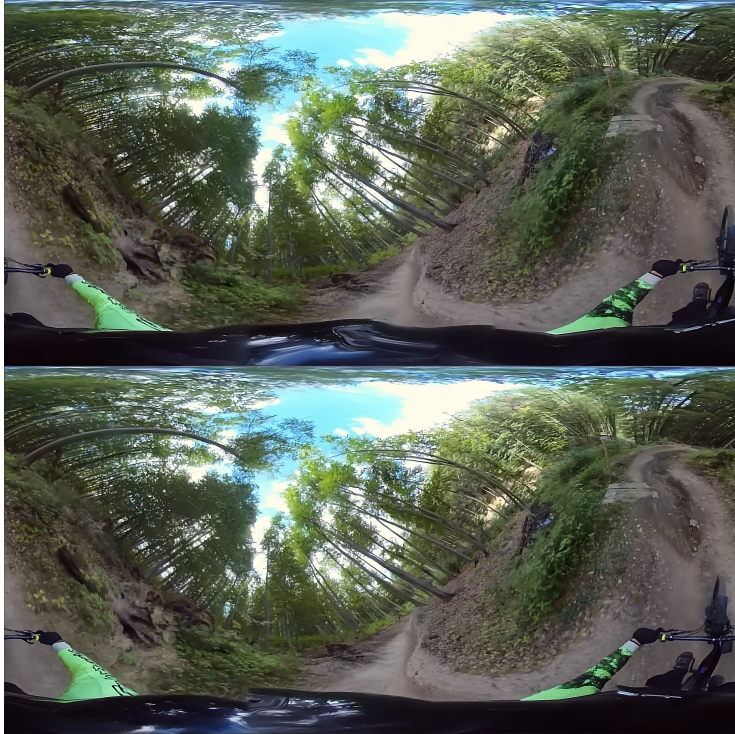

Fig. 2: The mountain scene
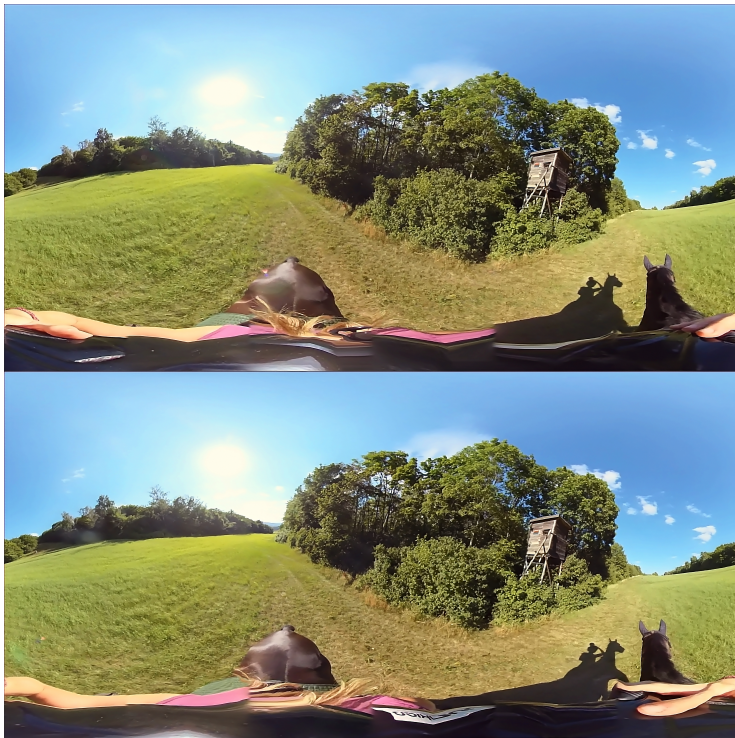
Fig. 3: The downhill scene

Fig. 4: The horse riding scene

Fig. 5: The climbing scene
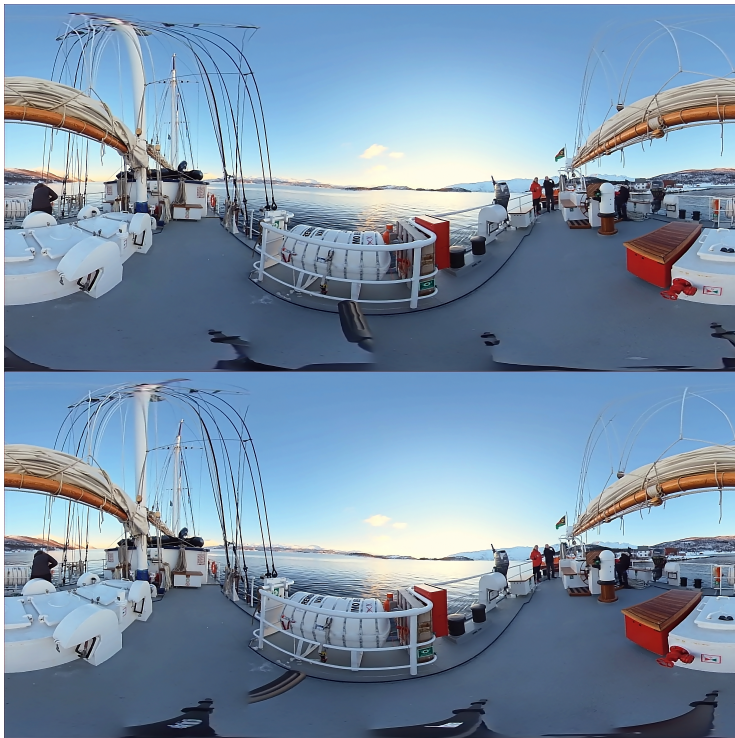

Fig. 6: The walking scene

Fig. 7: The cave scene

Fig. 8: The boat scene