Streaming of photo-realistic texture mapped on 3D surface

F. Jordan, S. Horbelt, T. Ebrahimi

Signal Processing Laboratory, Swiss Federal Institute of Technology,
CH 1015, Lausanne,
Switzerland
Email: jordan@ltssg3.epfl.ch, horbelt@desun1.epfl.ch, ebrahimi@epfl.ch

ABSTRACT

We present a novel technique for efficient coding of texture to be mapped on 3D landscapes. The technique enables to stream the data across the network using a back-channel. The use of Wavelet and Discrete Cosine Transforms is investigated and compared. This technology has been proposed has a tool for the emerging MPEG-4 standard.

1. Introduction

A new method to obtain an optimal coding of texture and 3D data is presented. It consists in taking into account the viewing position in the 3D virtual world in order to transmit only the *most visible* information. This approach allows to reduce greatly the amount of transmitted information between a remote database and a user, given that a back-channel is available (Figure 1). Only a fraction of the information is then sent, depending on object geometry and viewpoint displacement. This fraction is computed both at the encoder and at the decoder side. This basic principle is applied to the texture data to be mapped on a 3D grid mesh and is called "View-Dependent (VD) Scalability".

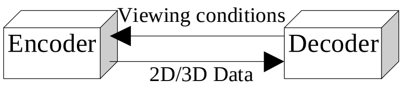


Figure 1: Usage of a Back-channel

The Section 2 explains why the texture image can be scaled, Sections 3 and 4 describe how Wavelet Transform (WLT) and Discrete Cosine Transform (DCT) can be used to perform this scaling. The technique is then applied for grid mesh data in Section 5. The Sctions 6 and 7 describe how DCT and WLT are used to stream and compress a texture image. Finally, results are presented in Section 8.

2. Principle of VD scalability

The VD scalability is based on the difference, which exists between a texture image and a texture mapped object. Two parameters are taken into account: the tilting of the plane with respect to viewing direction and the distance to the plane. It is possible to compute which amount L' of a texture image of size L is really visible (Figure 2) as shown by [2] in the case of a 1D-plane.

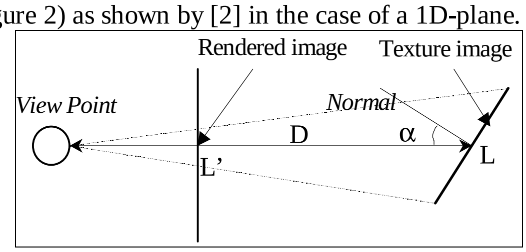


Figure 2: Effect of distance and tilting on rendered image.

This is generalized to the case of a 2D plane [3], by taking into account the rotation angle β of the plane (Figure 3).

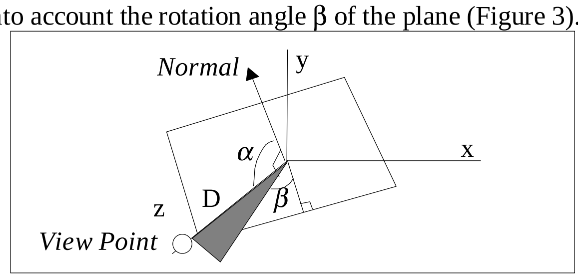


Figure 3: Generalization for a 2D-plane.

We have shown that all the information of the original texture is not visible on the rendered image. One way to select only the visible information is to filter the texture data. The filtering is done by setting to zero, or "shrinking" some of the coefficients in the transform domain.

3. Filtering using the tilting effect

3.1. Principle

In the following, it is shown how DCT and WLT can be used to filter the non-visible information due to tilting effect. This can be generalized for arbitrary β angles as shown in [5].

3.2. DCT

All DCT coefficients above the cut-off frequency cos(α), corresponding to the titling angle, are set to zero, where a value one stands for the highest frequency within a block.

3.3. WLT

The filtering is achieved by setting to zero the coefficients of some of the subbands. The number l of shrunk subbands can be written as: $l = -\dfrac{Ln(\cos(\alpha))}{Ln(2)}$

3.4. Comparison

Figure 4 gives an example of coefficients which are shrunk using the tilting effect. DCT technique scales linearly with cosα (while WLT scales with the Log of the Cosine). Moreover, the maximum shrinking (α close to 90°) will leave unchanged more than 1/4th of the coefficients for WLT and only 1/Nth of them for an NxN DCT.

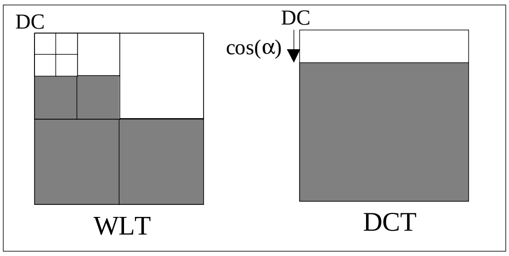


Figure 4: Filtering by tilting criterion α

4. Filtering using the distance effect

4.1. Principle

Assuming a perspective projection, when a texture mapped plane is far away from viewpoint, less information is visible. In the following it is shown how DCT and WLT can be used to low-pass filter this useless information.

4.2. DCT

All DCT coefficients which have one normalized frequency component higher than 1/D are set to zero.

4.3. WLT

The low-pass filtering operation is straightforward since the subband decomposition is available. Highest frequencies are set to 0 (HL, HH, LH). The number l of shrunk subbands is: $l = \frac{Ln(D)}{Ln(2)}$.

4.4. Comparison

The following figure shows which coefficients are shrunk, using the distance criteria. The scaling is again logarithmic for WLT and linear for DCT. The DCT technique has a scalability limited by the block size (N different values) which is not the case for WLT. Moreover, the wavelet representation is particularly suitable for graphic engines which use mip-mapping techniques.

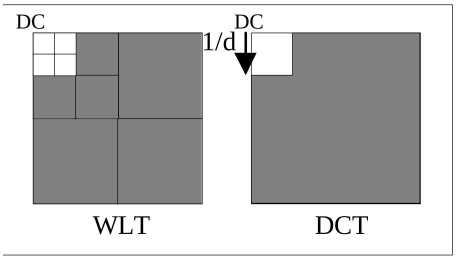


Figure 5: Filtering by distance criterion d

5. Application to grid mesh

5.1. Principle

In the case of a grid mesh on which a big texture is to be mapped, the same principle can be applied to each of the cells that define the mesh. Additionally, the texture blocks mapped on cells which are out of the field of view can be removed.

5.2. DCT

It is assumed that the number of cells and the texture size are such that a texture block of NxN pixels is mapped on each cell. The cell orientation influences only the DCT coefficients in the corresponding NxN block.

5.3. WLT

The size of the DC is equal to the size of the grid mesh. Coefficients are shrunk in each corresponding subband. For instance, if 8x8 blocks are used, 4x4 blocks are shrunk in the 3 highest subbands.

5.4. Comparison

The following figure shows which coefficients are modified. The DCT technique may require less memory than WLT since the shrinking is always bounded to one area of the transformed image (for a given cell).

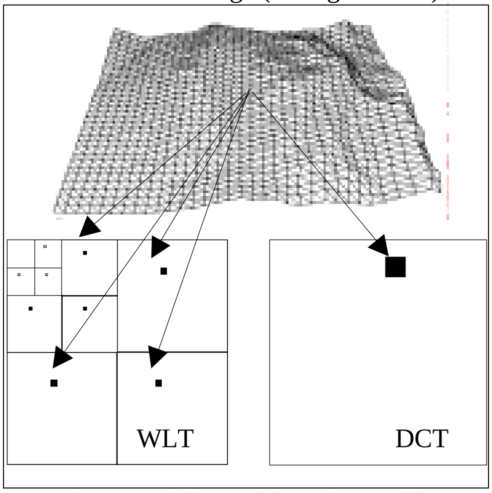


Figure 6: Influence of the position of one grid cell on the coefficients in the frequency domains.

6. Streaming

6.1. Principle

First the texture data needed for the first frame is transmitted, we call this data *intra data*. It is then straightforward to send only the new coefficients (*inter data*) as the observer is moving in the scene. Therefore, it is possible to update the texture information, according to the new viewer position using a very low bit rate channel. The streaming technique is totally identical for DCT and WLT transforms.

6.2. Comparison

For implementation purposes, a mask image is created which indicates which coefficient is to be transmitted.

The following figures show the difference between DCT and WLT mask images. For both techniques, it appears clearly that the Intra data size (Figure 7) is much higher compared to Inter data (Figure 8).

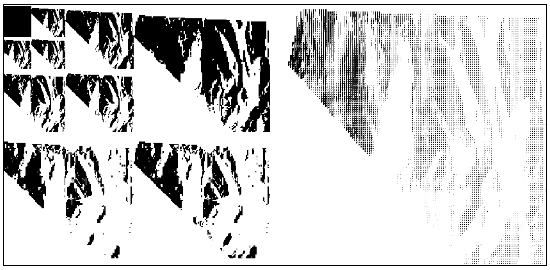


Figure 7: Example of Intra mask of transmitted coefficients for WLT (Left) and DCT (Right)

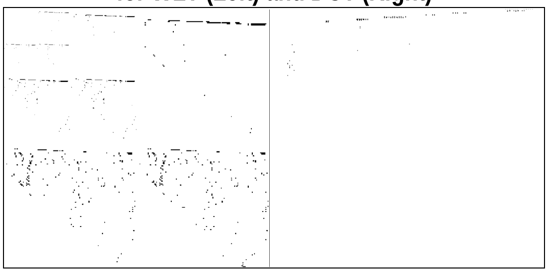


Figure 8: Examle of Inter mask of transmitted coefficients for WLT (Left) and DCT (Right)

7. Encoding

7.1. Principle

The encoding is performed in a conventional way similar to other still image compression techniques. Therefore, the syntax of the bitstream remains basically unchanged (Although the semantic is changing). The frequency transformation is made only once at the beginning of the session (DCT or WLT).

7.2. DCT

The DCT coefficient remaining after the shrinking are zigzag-scanned and coded using the quantization tables of MPEG-2 and the VLC tables of MPEG-4, (see [5]). The code had to be extended, because due to the shrinking and quantization, the inter blocks often contain only zero valued AC coefficients. So for each block a one bit entry in the Coded Pattern Value (CPV) table denotes whether any non-zero AC coefficients will be sent.

7.3. WLT

A regular wavelet encoder is used and basically nothing is changed neither in the encoding nor in the decoding process. Daubechies (9,3) biorthogonal filters are used. The decoded data is added to previously received coefficients prior to synthesis. In particular the zero-tree is always started from the root, whatever the mask is (which is sub-optimal). The DC is transmitted with the Intra data.

7.4. Comparison

With our current implementation, the DCT is more efficient than WLT since it takes the mask into account for encoding/decoding process.

8. Results

8.1. Introduction

The results are produced with a texture size of 1024 by 1024 texels (4:2:0), which represents 12 Mbits. The rendering screen size is 512 by 512 pixels. The test sequence consists of 76 frames. It simulates a realistic airplane flight over a landscape in Switzerland.
The bitrates and PSNR results are given in the Figure 9 and Figure 10.

8.2. DCT

Coder and decoder must store whether a coefficient has been already transmitted or not. This means one bit per coefficient or in our case 1 Mbit in total. The coder has the texture stored, quantized in frequency domain, so it only selects the coefficients and entropy codes them. The decoder knows for each Macroblock whether coefficients are expected. It decodes them, performs an inverse DCT and adds the results to the texture in the texture memory.

8.3. WLT

A buffer is also used to store which coefficient has already been transmitted (1 Mbit). The already sent coefficients are stored in a buffer of floating point numbers with the same size as original texture.

8.4. Comparison

The main difference comes from the way the texture is scaled in each case (linear for DCT and logarithmic for WLT). The intra frame is 472 kBits for WLT and 282 kBits for DCT, with a PSNR of 3 dB more for WLT. The Inter mean size is 5.7 kBits for WLT and 6.7 kBits for DCT. The total sum of transmitted bits is 10% higher in the case of WLT. The average PSNR value is roughly the same.

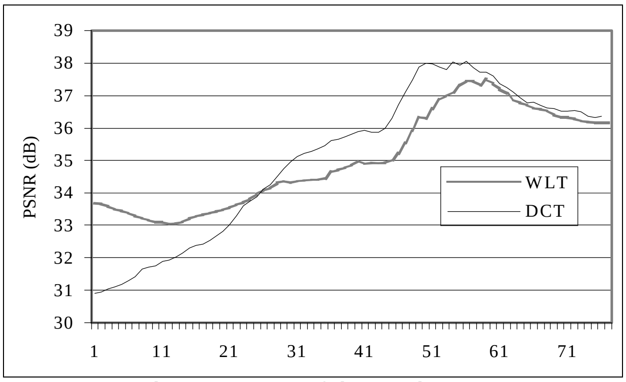


Figure 9: PSNR of the Luminance

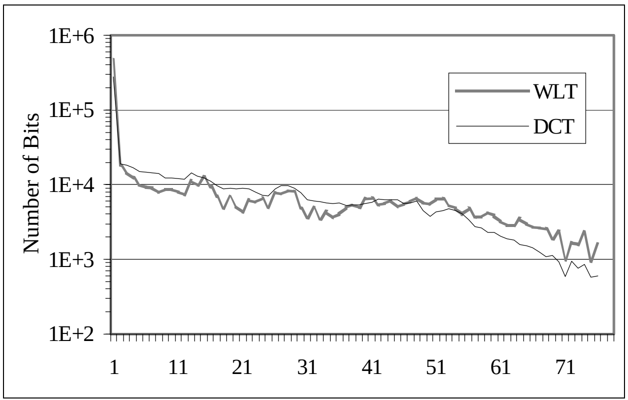


Figure 10: Number of bits sent for each frame of the test sequence using the WLT and DCT method.

9. Conclusion and perspectives

This technique has been proposed to ISO/MPEG-4 organization in the framework of Synthetic Natural and Hybrid Coding activities.

Several core experiments have been performed ([4][5][6][7]) which show that this technique performs better than the current Video Verification Model of MPEG-4 for the streaming of texture data.

Future work will merge and extend our experiments in order to stream bigger landscape areas and zoom over several magnitudes of resolutions, while keeping very-low bitrates. New View-dependent criteria should also be developed in order to take into account other viewing conditions such as speed, lighting, fog,.. etc.

Acknowledgments:

We would like to thank Swissair for providing us with photo-realistic landscape textures and elevation data. Special thanks also to Dr. Homer Chen from Rockwell who has actively contributed to the success of this work.

10. References

[1] Funkhouser, T.A. and Sequin, *Proceedings of SIGGRAPH 93*, pp. 247-254.

[2] Peter Lindstrom, David Koller, Larry F.Hodges, William Ribarsky, Nick Faust, “Level-of-detail management for real-time rendering of photo-textured terrain”, *Internal report of Georgia Tech*, http://www.cc.gatech.edu/gvu/research/techreports95.html

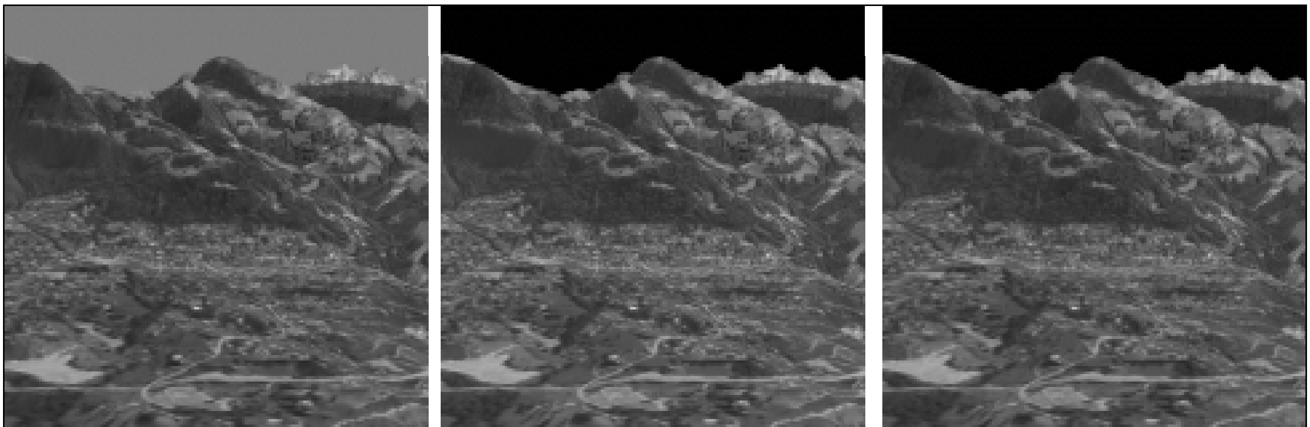
[3] Stefan Horbelt, Fred Jordan, Touradj Ebrahimi, “View-dependent texture coding for transmission of virtual environment”, *Proceedings of Conference on Image Processing and its Aplications*, July 1997, Dublin, Vol. 1, pp. 433-437.

[4] Fred Jordan, Touradj Ebrahimi, “View-dependent texture for transmission of virtual environment”, *MPEG96/M1501*, Maceio, 1996.

[5] Touradj Ebrahimi, Homer Chen, “Description of Core-experiment on visual texture/mesh coding”, *MPEG97/M2317* Stockholm, 1997.

[6] Stefan Horbelt, Fred Jordan, Touradj Ebrahimi, “Results of Core Experiment V1 using DCT”, *MPEG97/M2279*, Stockholm, 1997.

[7] Fred Jordan, Stefan Horbelt, Touradj Ebrahimi, “Results of Core Experiment V1 using Wavelets”, *MPEG97/M2278*, Stockholm, 1997.



WLT **Original** **DCT**

Figure 11: First frame of the sequence rendered with WLT (Left) and DCT (Right). The original is at the center.