

Scalable Video Coding Using Wavelets

Sai Karthik Vuppalapati (09d07050) Satya Naren (09d07051)
Swrangsar Basumatary (09d07040)
swrangsar@iitb.ac.in
Group 21

April 2013

Abstract

This is a report for the Scalable Video Coding Using Wavelets application assignment in the course EE 678 Wavelets. Our goal in this assignment is to demonstrate the spatial aspect of video scalability by encoding different spatial video resolutions into a single bitstream.

1 Introduction

Spatial Scalability within the H.264/AVC Scalable Video Coding Extension has already been demonstrated[2]. But here we are trying to demonstrate spatial scalability using wavelets because wavelets have the property of being inherently scalable.

In this experiment, we decided to use *three* different spatial resolutions for demonstration. We encoded the three different resolutions at the transmitter side into a single bitstream and transmitted it. At the receiver, we decoded the bitstream to get the resolution required.

2 Encoding the video

At first, we resized every frame of the video into a $2^N * 2^N$ square frame where,

```
N = floor(log2(max(size(frame))))
```

Then we did the following to each frame:

1. We did a 2-level 2-D wavelet decomposition to get the subbands $A2$, $H2$, $V2$, $D2$, $H1$, $V1$ and $D1$.
2. $A2$, the approximation subband, is our base layer. $H2$, $V2$ and $D2$ forms our first enhancement layer. The second enhancement layer is formed by $H1$, $V1$ and $D1$.
3. Then we compressed each layer using SPIHT[1] algorithm and quantized them.

Quantizing method

We quantized the compressed layers to achieve a peak Signal-to-Noise ratio (PSNR) that was above a user-specified minimum value.

3 Transmission

We then arranged every quantized frame so that for every frame the data was arranged in the following manner: the base layer came before the first enhancement layer and the first enhancement layer came before the second enhancement layer. Then we combined all the frames of the video into a single bitstream.

4 At the receiving end

At the receiving end, if we want the smallest of the three pre-decided resolutions then we retain only the

base layer data of each frame. And if we want the intermediate resolution then we retain the first enhancement layers also. For the maximum resolution, we retain all the layers.

We take the retained data and then decode them to get the output.

Decoding the video

The decoder takes in the layers of each frame and uncompress them using SPIHT[1] algorithm. The uncompressed layers are combined using 2-D wavelet reconstruction and then resized to a user-specified dimension.

These resized frames together form the output video.

References

- [1] Amir Said and William A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3), 1996.
- [2] C. Andrew Segall and Gary J. Sullivan. Spatial scalability within the h.264/avc scalable video coding extension. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9), September 2007.