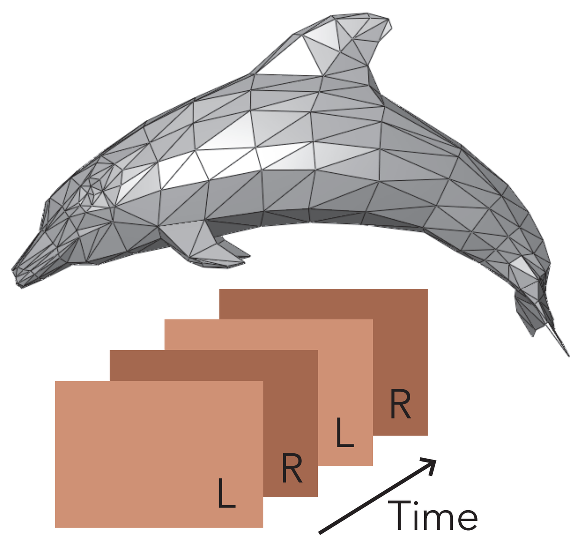
Compression and Transmission of Multiview Video Data

Ricardo da Silva
Supervisors: Rong Zhang, Lajos Hanzo
University of Southampton

3D displays are a relatively novel technology, and there are many different ways of creating a 3D viewing experience. This project investigated compression and transmission schemes for multi-view and stereoscopic video, and presents implementation results of simple motion compensation and DPCM codecs.

Video representation → Compression and transmission → Display



MVC
Multiview video coding

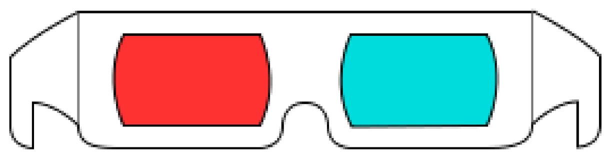
HEVC
High efficiency video coding

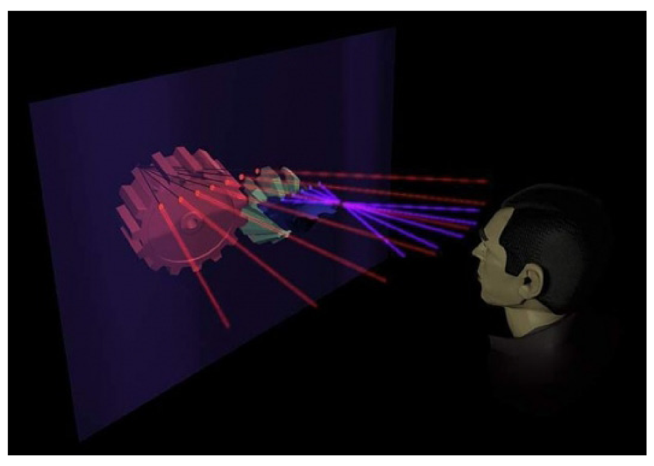
MRSC
Mixed resolution stereo coding [1]

Turbo codes • Reed Solomon • UEP [2]

Objective measure of reconstructed frame quality:

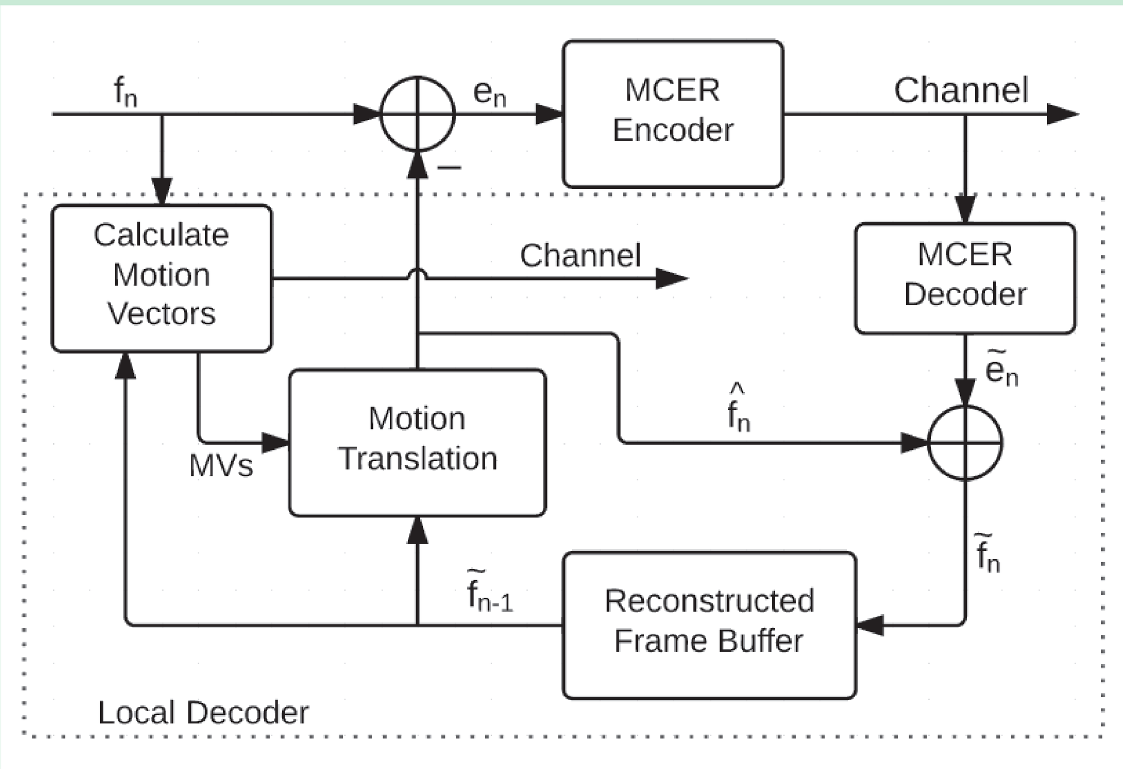
$$PSNR = 10\log_{10}\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\max^2}{\sum_{i=1}^{p}\sum_{j=1}^{q}\left(f_n(i,j)-\tilde{f}_n(i,j)\right)^2}$$



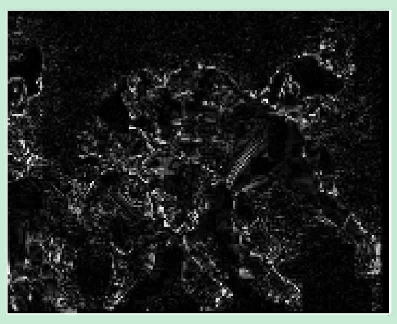


Motion compensation

Motion Compensation compresses the energy of the signal to be transmitted, by predicting blocks of pixels in terms of the linear movement of a corrresponding block of pixels in the previous frame. [3]

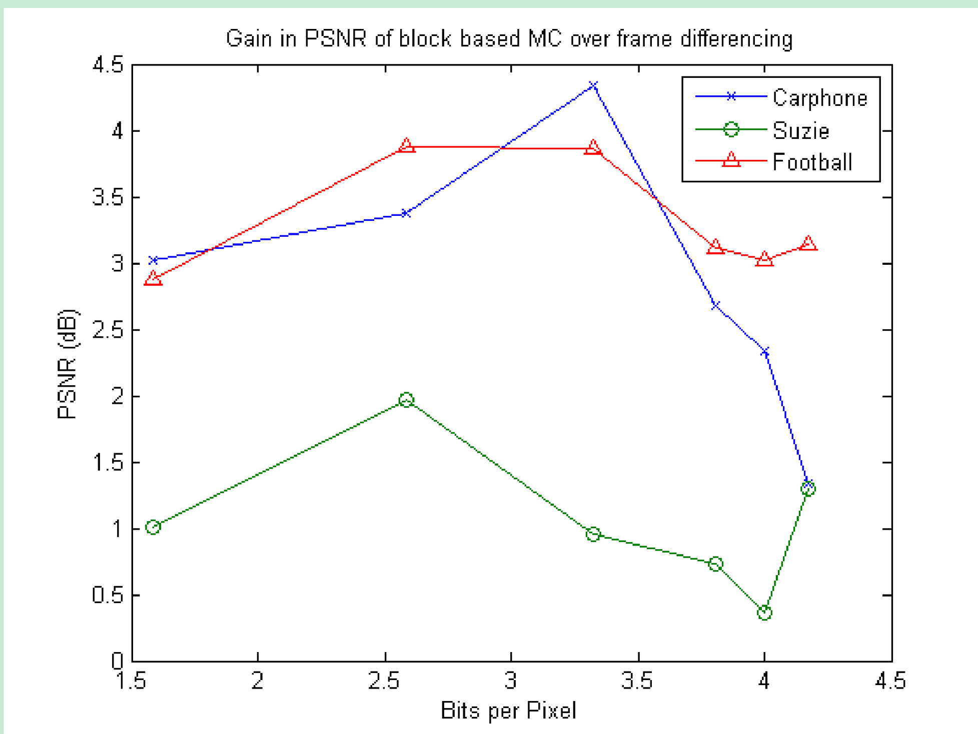


Difference between original and reconstructed frame using only the motion vectors (MCER set to 0):

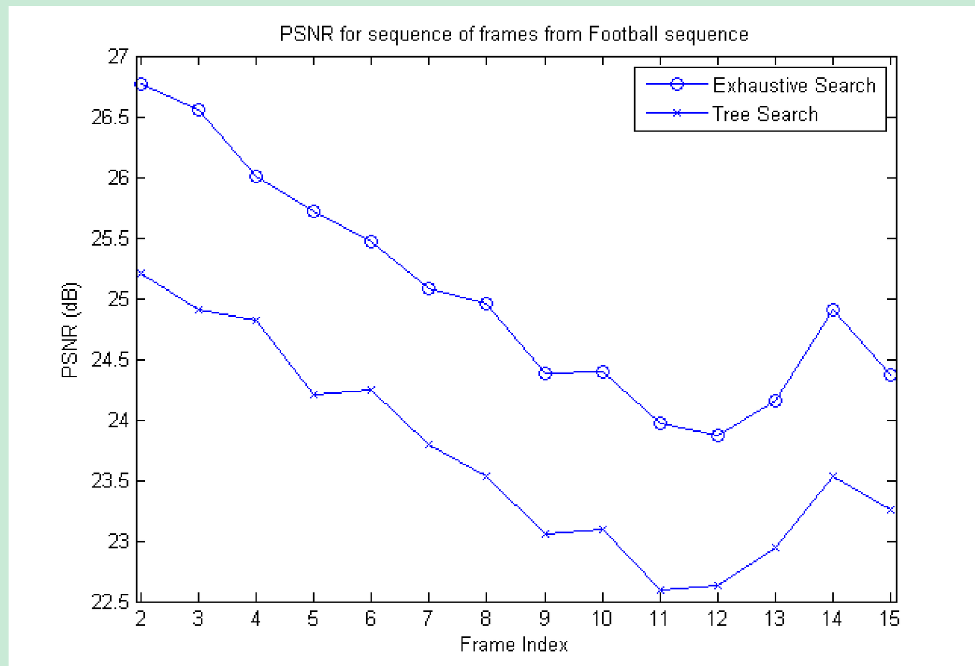


Frame differencing is the simplest form of motion compensation, but is not very useful for realistic video sequences. A block based motion estimation codec is designed, which is shown to work better than frame differencing, especially for video sequences with high amounts of motion. Block and search window sizes of 8 and 16 pixels, respectively, were used in this implementation, and Max-LLoyd quantisers were trained for each video sequence.

Football and Carphone have more movement than Suzie, and hence block motion compensation outperforms frame differencing in these videos.



An iterative tree based motion search algorithm was shown to be far less complex, with only a small reduction in quality, when compared with an exhaustive search. This converges in only 27 block comparisons, rather than the 256 comparisons required by a full search.



The Minimum Mean Squared Error was used to select the optimum block in the previous frame, for both tree and full search.

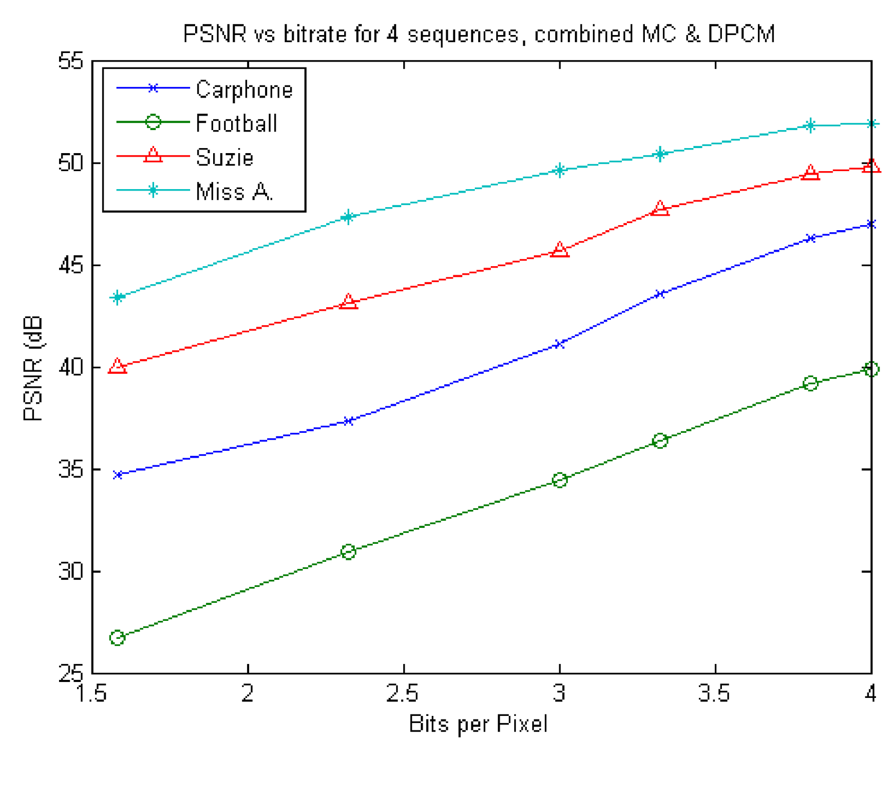
$$M_{mse} = \arg\min_{m_x,m_y}\sqrt{\sum_{i=1}^{b}\sum_{j=1}^{b}\left(f_n(x+i,y+j)-f_{n-1}(x+i-m_x,y+j-m_y)\right)^2}$$

DPCM

Differential pulse code modulation is based on the principle that adjacent pixels in an image are likely to be similar, and hence redundancy can be removed by representing each pixel in terms of an adjacent (usually previous) pixel.

$$e_n = i_n - a \cdot i_{n-1}$$

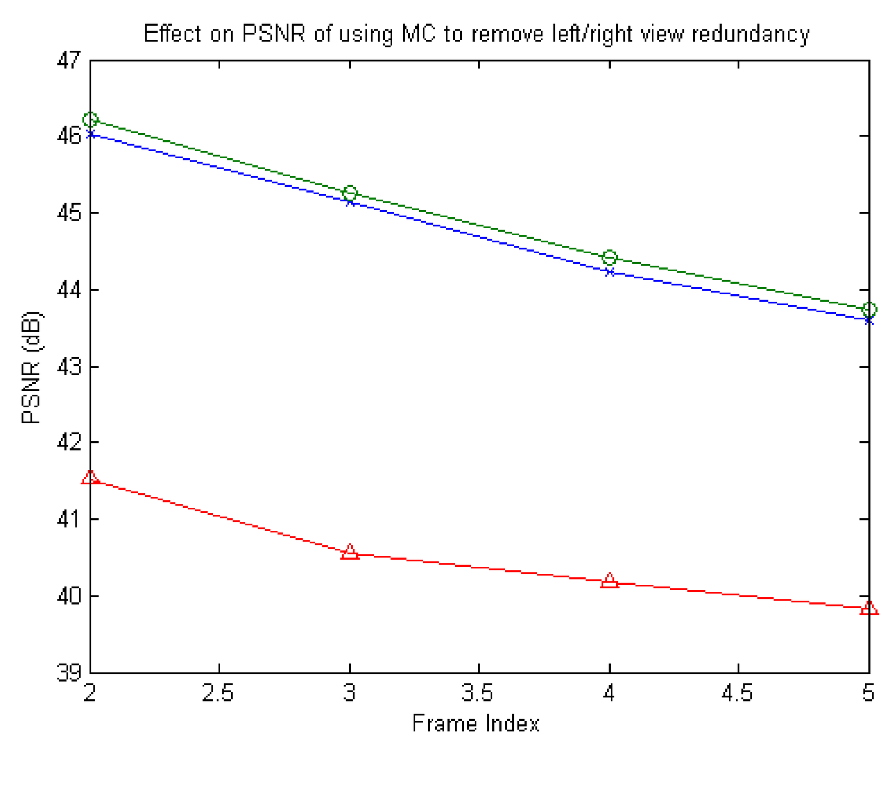
A DPCM codec was implemented and paired with block based motion compensation, which further reduced the energy of signals to be transmitted, resulting in good overall PSNR vs bitrate values.



Application to Stereoscopic Video

Removing redundancy between left/right frames: complex methods exist, but how do simple methods fare? The block based motion compensation algorithm was used to create a representation of the right view in terms of movement in the left view, which resulting in a lower energy representation of a single frame.

However, this introduced noise that negatively impacted further compression. It is better to encode the left/right sequences individually (Simulcast), or use more complex prediction [4].



Conclusions

State of the art three dimensional viewing systems were reviewed, along with the relevant video representations and compression techniques. In order to build understanding of the fundamentals of these high complexity codecs, a simple motion compensation predictive codec was implemented, as well as a differential pulse code modulation codec.

The performance of these two systems was investigated, and results were presented that showed how they improve the PSNR for a given bitrate. Stereoscopic video was also tested with the same algorithms. It was found that the simple motion compensation algorithm is not suited to predicting right from left frames in stereoscopic video, as the resulting prediction is too noisy.

References

[1] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 1–4, May 2009.

[2] A. Gotchev, G. Akar, T. Capin, D. Strohmeier, and A. Boev, "Three-dimensional media for mobile devices," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 708–741, April 2011.

[3] L. L. Hanzo, P. Cherriman, and J. Streit, *Video compression and communications*. John Wiley & Sons, 2007.

[4] N. Ramzan, A. Amira, and C. Grecos, "Efficient transmission of multiview video over unreliable channels," *Image Processing (ICIP), 2013 20th IEEE International Conference*, pp. 1885–1889, Sept 2013.