

Structural Similarity Based Efficient Multi-View Video Coding

L.C.Manikandan¹, M.A.Anusha², Dr. A. Lenin Fred³

¹(Asst.Professor) Department of Computer Science and Engineering, Mar Ephraem College of Engineering and Technology, Marthandan, Tamilnadu, India.

²(PG Student) Department of Computer Science and Engineering, Mar Ephraem College of Engineering and Technology, Marthandam, Tamilnadu, India.

³(Professor) Department of Computer Science and Engineering, Mar Ephraem College of Engineering and Technology, Marthandam, Tamilnadu, India.

Abstract: - Efficient compression of multi-view images and video is crucial for 3DV application. The major task of three-dimensional video (3DV) coding is to provide high quality depth data. For experimental coding purposes, the compression method used is H.264/AVC video coding standard. In this paper, Structural similarity (SSIM) based Rate distortion (RD) optimization is expressed for distortion metric. In multi-view video, the left view and the right view have their pixels with strong dependencies of structural information that are spatially close. The full reference quality metric SSIM is used to exploit the structural information from different views in multi-view video coding. With the proposed system, significant gains can be achieved in terms of SSIM index.

Keywords: - DCT, MB, Multi-view video coding, SSIM, Rate distortion optimization

I. INTRODUCTION

Today's advances in display and camera technology enable new applications for three-dimensional video (3DV) [1], [2] and have become one of the promising fields regarding the development of new application for natural scenes. The 3-D video proffers high quality multimedia experience to the consumer through display, signal and transmission technologies. The display and receiver side of 3-D video may gain an increased level in the production of 3-D content [3]. For the 3DV, the display technology may need multi-views and for that, at least two views are necessary.

With the High efficiency video coding (HEVC) based multi-view compression scheme [4], forward compatibility with HEVC is guaranteed similarly to the forward compatibility provided by Multi-view video coding (MVC). The MVC was enabling a certain view to be inter-view predicted from an earlier decoded view. The encoder can choose a Rate Distortion (RD) optimal way in the interview prediction. Additional to this, the complexity restriction limiting inter-view prediction is only for within the same time instance. In the proposed system, the full reference quality metric SSIM is used to exploit the structural information from different views in multi-view video coding. The rest of this paper is organized as follows. Section II gives [5], [2] an introduction to multi-view video coding and the compression techniques for multiple views. Section III will show an introduction to structural similarity, Section IV [4], [6], [7] will discuss the proposed changes with structural similarity to enable multi-view video coding.

II. MULTIVIEW VIDEO CODING

The multiple video cameras are used to simultaneously acquire various view points of the scene. The resulting data are often referred to as multi-view video [5]. As the multi-view video imagery captures the same 3-D scene from different viewpoints, there exist inter-view statistical dependencies among the images [2]. The similarities between the images can be classified into Inter-view similarity and Temporal similarity,

2.1 Inter-view similarity

Objects in each image are subject to parallax and appear at different pixel locations. To exploit this inter-view similarity, disparity compensation techniques are used.

2.2 Temporal similarity

Objects appear in successive images but at different pixel locations. To exploit these temporal similarities, motion compensation techniques have been developed.

III. STRUCTURAL SIMILARITY

The structural similarity (SSIM) index is a method for measuring the similarity between two images/videos. The SSIM index is a full reference metric, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve on traditional methods

like peak signal-to-noise ratio (PSNR) and mean squared error (MSE). The MSE or PSNR approaches estimate perceived errors, SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene.

The SSIM metric is calculated on various windows of an image. The measure between two windows x and y of common size $N*N$ is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where,

μ_x average of x , μ_y average of y , σ_x^2 variance of x , σ_y^2 variance of y , σ_{xy} covariance of x and y , $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are the two variables to stabilize the division with weak denominator, L the dynamic range of the pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$ by default. In the structural similarity index, the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image and compression quality. Here the MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represent a measure of peak error.

IV. STRUCTURAL SIMILARITY BASED MULTIVIEW VIDEO CODING

In multi-view video coding, the left view and the right view are taken separately, and then the Picture Order Count (POC) will provide a counter that specifies the relative order of the picture in the bit stream, in output order. The relative order of the picture is indicated in POC, rather than the timing of the picture. This allows the systems to carry the video bit stream, to control the exact timing of the processing and output of the video bit stream, without affecting the decoding process for the values of the samples in the luma and Chroma sample arrays of the picture.

The Reference Parameter Set (RPS) will set the reference frames to the left view and the right view based upon the similarities between the views. The RPS in H.264/AVC [4], enable motion compensation with other pictures, a buffer with available reference pictures called the Decoded Picture Buffer (DPB) is maintained. The Figure 4.1 shows the architecture for structural similarity based multi-view video coding. The Inter-view prediction process may be applied between the different views in the same time instant. The inter-view predictions with frames from a different time instance are not allowed. The SSIM based RD optimization [6], [7] is an image quality metric in image/video processing due to its good compromise between quality evaluation accuracy and computation efficiency. The SSIM index may improve the performance and quality in the visual scene. Finally the compressed video data is found and then the decoder will decode the data.

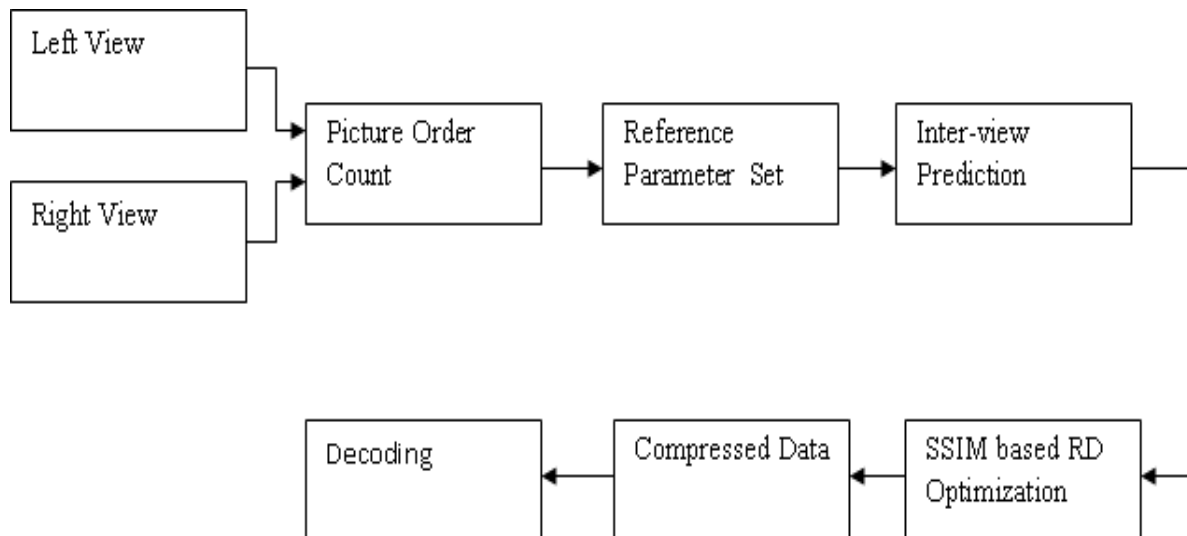


Figure 4.1 structural similarity based multi-view video coding

V. EXPERIMENTAL RESULTS

For the compression point of view, in multi-view video coding we need left view and right view separately. Here Figure 5.1 (a) and (b) shows the left and right views of the video.

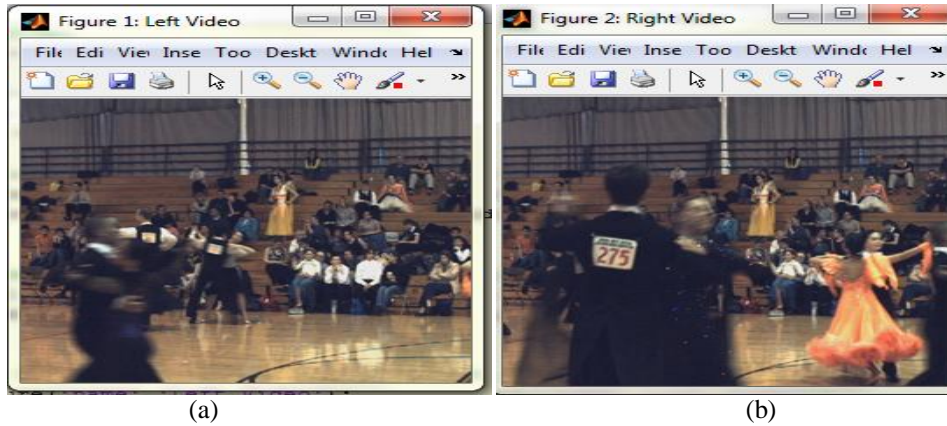


Figure 5.1(a) left view of the video (b) right view of the video

The basic concept of inter-view prediction for MVC is to exploit both spatial and temporal redundancy for compression. Inter-view redundancy may occur due to the cameras with multi-view videos from nearer views. So for the prediction process, inter-view prediction may be applied between the two views at the same time instant [3]. For the left view and right view, their pixels have strong dependencies with structural information of a scene, to extract the structural information and to improve the quality of a video, we propose a new method in this paper i.e. structural similarity (SSIM) index. Figure 5.2 shows the interview and SSIM image of the video. The Similarity measurement [10] of two image signal x and y in the system takes the combination of three components - luminance, contrast and structure:

$$S(x, y) = [l(x, y)] [c(x, y)] [s(x, y)]$$

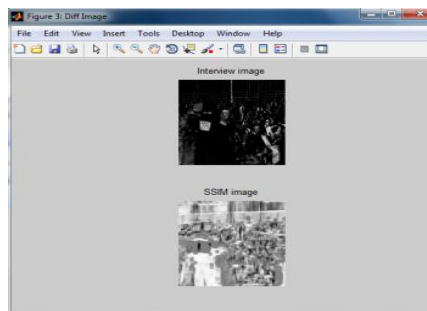


Figure 5.2 interview and SSIM of the video

The decoder may be stated that the dependencies would restrict the decoder to sequential decoding of every frame in different views [4]. For the decoding process, only the first two frames are to be decoded sequentially. The Figure 5.3 shows the decoded 3D video.

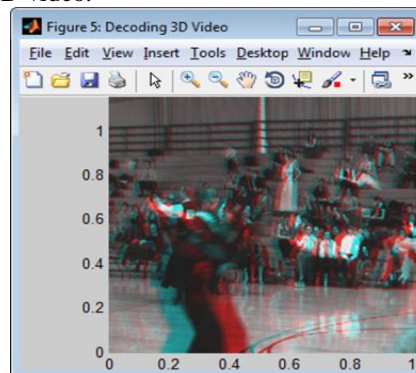


Figure 5.3 decoded 3D video

Distortion is the alteration of the original shape of something, such as, an object, an image or a video. While video coding, to get the quality of the video, we make the minimum distortion level for the video data. Figure 5.4 shows the minimum distortion level for the proposed system.

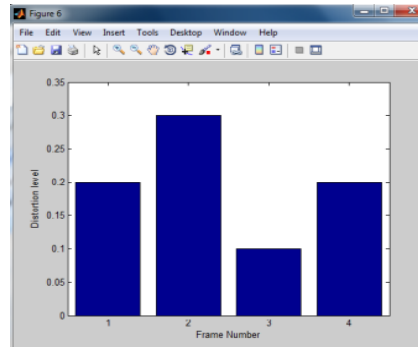


Figure 5.4 frame with distortion level

The peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible value of a signal and the power of distorting noise that affects the quality of its representation. Because, many signals have a very wide dynamic range, (ratio between the largest and smallest possible values of a changeable quantity) the PSNR is usually expressed in terms of the logarithmic decibel scale. In PSNR, we take the square of the peak value in the image (in case of an 8 bit image, the peak value is 255) and divide it by the mean square error. The compression ratio is the ratio between the maximum to minimum volume of the video. For streaming audio and video, the compression ratio is defined in terms of uncompressed and compressed bit rates instead of data sizes. The compression ratio is calculated as,

$$\text{Compression ratio} = \frac{\text{Uncompressed size}}{\text{Compressed size}}$$

The Figure 5.5 and Figure 5.6 show the improved PSNR value and compression ratio for the proposed SSIM index.

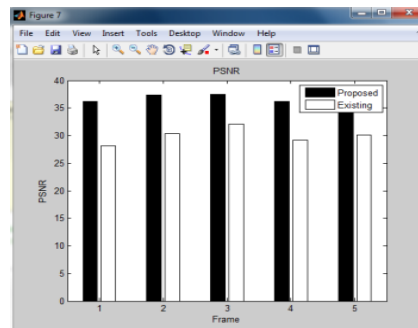


Figure 5.5 PSNR value

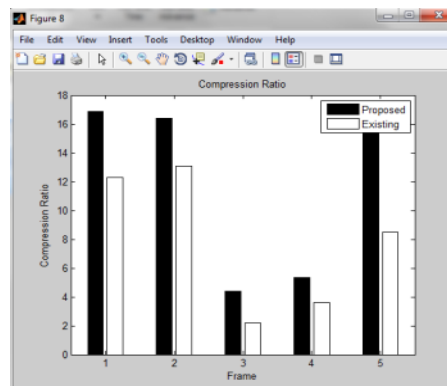


Figure 5.6 compression Ratio

VI. CONCLUSION

In this paper, a multi-view compression algorithm based on structural similarity (SSIM) index and a solution is forward compatible with H.264/AVC is proposed. The SSIM will exploit the structural information from the multiple views in multi-view video coding. With the purely SSIM based multi-view video coding, the quality of the video may be improved with the increased level of PSNR value and compression ratio. Furthermore, the proposed system achieves a significant gain in SSIM performance and in the quality of visual scene.

REFERENCES

- [1] M. Flierl, A. Mavlankar, and B. Girod, "Motion and Disparity Compensated Coding for Multi-view Video", vol.17, no.11, pp.1474-1484, *IEEE Trans. Circuits Syst. Video Technol.*, November 2007.
- [2] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient Prediction Structures for Multi-view Video Coding", vol.17, no.11, pp.1461-1473, *IEEE Trans Circuits Syst. Video Technol.*, November 2007.
- [3] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the Stereo and Multi-view Video Coding Extensions of the H.264/MPEG-4 AVC Standard", vol. 99, issue 4, pp.626-642, *Proceedings of the IEEE, Special Issue on "3D Media and Displays"*, April. 2011.
- [4] G.VanWallendael, S.Van Leuven, Jan De Cock, F. Bruls, and R.Van deWalle, "3d Video Compression Based on High Efficiency Video Coding", vol.58, No.1, pp.137-145, *IEEE Transactions on Consumer Electronics*, February 2012.
- [5] M.Flierl, and B. Girod, "Multi-View Video Compression", pp.1-21, *IEEE Signal Processing Magazine*, July 2007.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility To Structural Similarity", vol. 13, no. 4, pp. 600–612, *IEEE Transactions on Image Processing*, April 2004.
- [7] S.Wang, A.Rehman, Z.Wang, Siwei Ma, W.Gao, "Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization", vol.22,no.4, pp.1418-1429, *IEEE Transactions on Image Processing*, April 2013.
- [8] K. Müller, P. Merkle, and T. Wiegand, "3-D Video Representation Using Depth Maps", vol.99, no.4, pp.643-656, *Proceedings of the IEEE*, April 2011.
- [9] C.Yeo, H.Li.Tan, and Y.H.Tan, "On Rate Distortion Optimization Using SSIM", pp.833-836, *International Conference on Acoustics, Speech, and Signal Processing*, 2012.
- [10] Ming-Jun Chen, and Alan C.Bovik, "Fast Structural Similarity Index Algorithm", pp.994-997, *International Conference on Acoustics, Speech, and Signal Processing*, 2010.