Enhanced Depth Video Coding for 3D videos

Shivam S. Agrawal[1], Amol S. Bendmali[2]

[1]Assistant Professor, Department of E&TC, PLITMS, Maharashtra, India
[2]Assistant Professor, Department of E&TC, PLITMS, Maharashtra, India

ABSTRACT

With the recent development of 3D multimedia/display technology and the increasing demand for realistic multimedia, 3D video has gained more attention as one of the most dominant video formats with a variety of application such as 3-dimentional TV or Free view point. In 3D video processing, depth video coding is very essential. In depth coding, the object boundaries are hard to compress and severely affect the rendering quality since they are sensitive to coding errors. A depth boundary reconstruction filters is used and utilize it as an in-loop filter to code the depth video. The proposed depth boundary reconstruction filter is useful for efficient depth coding as well as high-quality rendering.

Keyword:- *3D Video1, Depth Map2, Boundary Reconstruction3, stereoscopic4, Multimedia5, Video Processing6.*

1. INTRODUCTION

With the development of multimedia technology and the increasing desire for realistic media, there have been several studies on three-dimensional imagery. A stereoscopic image consisting of left and right images is able to show a realistic scene using special stereo displays. Several types of stereoscopic displays have been developed and most require viewers to wear glasses to view the 3D scene. Even though stereoscopic images provide an impressive 3Dexperience, there exist further challenges to view a scene from multiple 3D viewpoints and support auto stereoscopic displays. Auto stereoscopic displays provide highly realistic 3D images and free-view navigation to viewers without the need to wear glasses; this is achieved by generating various viewpoint of the scene.

As 3D video requires multiple video sequences captured from different camera positions, it becomes challenging to transmit and store such a large amount of data. This has lead to significant interest in investigating efficient view rendering methods. For this purpose it has been studied how to sample the plenoptic function or the light field to reconstruct missing views using image based rendering techniques. To improve the rendered view quality, multi view video plus depth formats are being developed, where only selected views are coded along with their corresponding depth maps, and other views are interpolated at the decoder using depth image based rendering (DIBR) technique.

2. METHODOLOGY

Unlike common color images, depth images exhibit a high degree of spatial correlation except at object boundaries. Thus, the assumption is that the complex regions such as object boundaries are more sensitive to coding errors and need much more coding bits compared to flat regions such as backgrounds. There are three algorithms to generate the depth videos the names of the algorithms are Laplacian of Gaussian Filter, Sobel operator and the Boundary Reconstruction Filter. The detail information of these all three algorithms is as below

2.1 LAPLACIAN OF GAUSSIAN FILTER

The Laplacian is a 2D isotropic measure of the 2nd order derivative of an image. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for differentiation of edges. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise, and hence the two variants individually will be described together here.

Initially the work will proceed with the Laplacian kernels. Laplacian kernels usually are the derived values of the second order derivative which can be taken as per the specification, the higher the kernel values the greater would be its enhancement Since the input image is represented as a set of discrete pixels and find a discrete convolution kernel that can approximate the second derivatives in the definition of the Laplacian.

Laplacian filters are derivative filters used to find areas of rapid change (edges) in images. Since derivative filters are very sensitive to noise, it is common to smooth the image (e.g., using a Gaussian filter) before applying the Laplacian. This two-step process is call the Laplacian of Gaussian (LoG) operation. The equation for the Laplacian of Gaussian function is

$$\mathrm{LoG}\,(x,y) = \frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] e^{\frac{-(x^2+y^2)}{2\sigma^2}} \qquad (1)$$

Where x, y takes the horizontal and vertical pixel notations, represents the standard deviation and is a constant.

When using the filter given above, or any other similar filter, the output can contain values that are quite large and may be negative, so it is important to use an image type that supports negatives and a large range, and then scale the output. Alternatively, a scaling factor can be used on the filter to restrict the range of values.

2.2 SOBEL OPERATOR

The Sobel operator, sometimes called the Sobel–Feldman operator or Sobel filter, is used in image processing and computer vision, particularly within edge detection algorithms where it creates an image emphasizing edges. It is named after Irwin Sobel and Gary Feldman, colleagues at the Stanford Artificial Intelligence Laboratory (SAIL). It was co-developed with Gary Feldman at SAIL. Sobel and Feldman presented the idea of an "Isotropic 3x3 Image Gradient Operator" at a talk at SAIL in 1968. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel–Feldman operator is either the corresponding gradient vector or the norm of this vector.

The Sobel–Feldman operator is based on convolving the image with a small, separable, and integer valued filter in the horizontal and vertical directions and is therefore relatively inexpensive in terms of computations. On the other hand, the gradient approximation that it produces is relatively crude, in particular for high frequency variations in the image.

The Sobel operator is slower to compute than the LoG, but its larger convolution kernel smooths the input image to a greater extent and so makes the operator less sensitive to noise. The operator also generally produces considerably higher output values for similar edges, compared with the LoG.

The operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives one for horizontal changes, and one for vertical. If we define A as the source image, and G*x*and G*y*are two images which at each point contain the horizontal and vertical derivative approximations respectively, the computations are as follows

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * A \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * A \qquad (2)$$

Where here denotes the 2 dimensional signal processing convolution operation. Since the Sobel kernels can be decomposed as the products of an averaging and a differentiation kernel, they compute the gradient with smoothing.

2.3 BOUNDARY RECONSTRUCTION FILTER

Boundary reconstruction filter contain main tree parameter: 1) occurrence frequency 2) similarity and 3) closeness. The occurrence frequency of the pixel represents the comparison between the current frequency pixel values to the neighbored frequency pixel value for an every pixel of the image and find the best intensity value which as maximum cost value. That is, the current pixel is replaced with the best intensity value same operation is perform for the similarities and closeness of the pixel difference is that in similarities it’s finds the similarities between the original pixel to the neighbored pixel to form a new pixel intensity value.

The depth boundary reconstruction filter is the combination of three parameters such as occurrence frequency, similarities and closeness as shown below

$$J_{recon}(k) = J_F(k) + J_S(K) + J_c(K) \qquad (3)$$

Where K represent the pixel intensity values. The two main advantage of the boundary reconstruction filter over others filters is it is more robust against outliers; a single pixel will not affect filtering significantly and since the filtered value must actually be the value one of the pixels in the neighborhood, filtering does not create new unrealistic pixel values.

The results show that the boundary reconstruction filter gives the much better result than the other two filters. These filtered depth video can be used for the boundary reconstruction filter to form the best realistic 3D videos.

2.4 EVALUATION OF DEPTH CODING

The coding efficiency of the image is measure in rate and distortion matrix. The PSNR values are used as the rate and distortion measures. PSNR is abbreviation of Peak Signal to Noise Ratio. PSNR is easily calculated and thus it is widely used for measuring the quality of reconstructed image. Usually high PSNR represents high quality of reconstructed image while low PSNR represents low quality of reconstructed image. Logarithmic scale is used for measuring PSNR. It is defined as ratio of maximum possible power of image data and power of noise is the error introduced by compression that affects the fidelity of its representation. It is calculated as follow.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{Max_I}^2}{\mathrm{MSE}}\right) \qquad (4)$$

3. RESULTS

The objective of this research is to develop a new 3D video coding system which can provide better coding efficiency with improved subjective quality as compared to existing 3D video systems such as the depth image based rendering (DIBR) system. The comparison graphs in between the PSNR and the no. of frames and three methods are to compare is LoG (Laplacian of Gaussian Filter), Sobel operator and Boundary Reconstruction Filter for the input sequence. The input sequences are video1_wildlife.wmv. The comparison graphs as shown below.

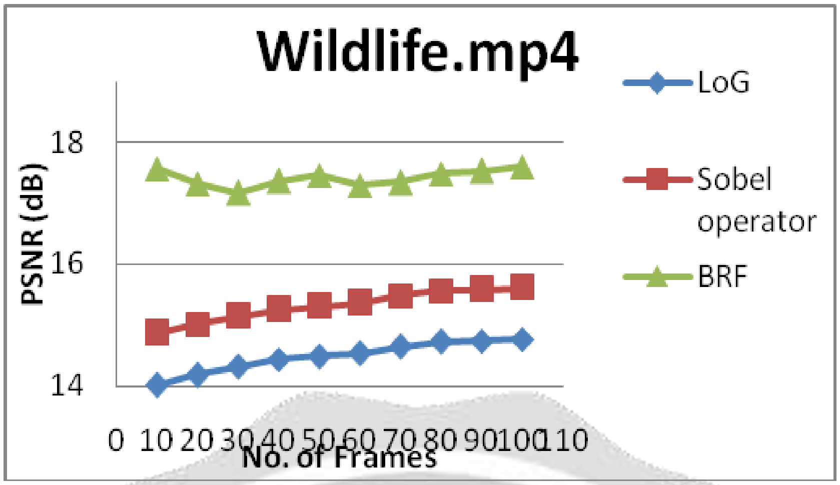

Chart -1: The comparison graph of Depth Video Coding

In addition, the results also show that the spatial down/upsampling approach outperformed the depth dynamic range reduction approach in both test sequences.

4. CONCLUSION

Use Boundary Reconstruction Filter is an effective technique in which 3D video depth video technique which recovers depth video without compromising on quality. PSNR calculated is approximately equals to 7.00 for boundary reconstruction method which show that videos are recovered with high fidelity.

Boundary Reconstruction Filter method originated from occurrence frequency similarities and closeness of the pixel, and was optimized by utilizing the unique characteristics of depth video: smooth depth level distribution within an object and sharp depth level variation near object boundaries.

5. REFERENCES

[1] P. Merkle, Y. Morvan, A. Smolic, K. Műller, P.H.N. de With, and T. Wiegand, "The effect of depth compression on multiview rendering quality," 3DTV Conference 2008, pp. 245–248, May 2008.
[2] W. S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view," in Proc. SPIE VisualInformation Processing and Communication (VIPC 2010), vol. 7543, p. 75430B, Jan. 2010.
[3] K.J. Oh, S. Yea, A. Vetro, and Y.S. Ho, "Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3-D Video," IEEE Signal Processing Letters, vol. 16, no. 9, pp. 747–750, Sept. 2009.
[4] Q. Zhang, P. An, Y. Zuo and Z. Zhang, ‘‘ A novel depth coding scheme for multiview video plus depth compression’’ in proc. of ICALIP, pp. 1169-1174, 2010.
[5] S.U. Yoon and Y.S. Ho, "Multiple color and depth video coding using a hierarchical representation," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1450–1460, Nov. 2007.
[6] Kwan-Jung oh, Yo-Sung Ho, "Depth coding using boundary reconstruction filter for 3D video system," IEEE Transactions on Circuits and Systems for Video Technology, pp. 1-10, 2011.
[7] S.T. Na, K.J. Oh, and Y.S. Ho, "Joint coding of multi-view video and corresponding depth map," International Conference on Image Processing (ICIP 2008), pp. 2468–2471, Oct. 2008.
[8] S.Y. Kim and Y.S. Ho, "Mesh-based depth coding for 3D video using hierarchical decomposition of depth maps," International Conference of Image Processing (ICIP), vol. 5, pp. 117–120, Oct. 2007.
[9] K. Müller, P. Merkle, and T. Wiegand, "3-D Video representation using depth maps," Proceeding of the IEEE, vol. 99 , No.4, pp. 643-656, April 2011.
[10] S.U. Yoon and Y.S. Ho, "Multiple color and depth video coding using a hierarchical representation," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1450–1460, Nov. 2007.

[11] H. Oh and Y.S. Ho, "H.264-based depth map sequence coding using motion information of corresponding texture video," Pacific RimSymposium on Advances in Image and Video Technology (PSIVT '06), pp. 898-907, Hsinchu, Taiwan, Dec. 2006.
[12] P. Merkle, Y. Morvan, A. Smolic, K. Műller, P.H.N. de With, and T. Wiegand, "The effect of depth compression on multiview rendering quality," 3DTV Conference 2008, pp. 245–248, May 2008.
[13] W. Matusik and H. Pfister, "3-D TV: A scalable system for real-time acquistion, transmission and autostereoscopic display of dynamic scenes," ACM Trans. Graph., vol. 23, no. 3, pp. 814–824, Aug. 2004.
[14] M. Tanimoto, "Overview of free viewpoint television," Signal Processing: Image Communication, vol. 21, pp. 454–461, July 2006.
[15] K.J. Oh, S. Yea, and Y.S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-D video," Picture Coding Symposium (PCS), May 2009.
[16] M. K. Kang, C. Lee, J. Y. Lee, and Y. S. Ho, "Object adaptive depth compensated inter prediction for depth video coding in 3d video system," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME2010), pp. 1230–1235, Jul. 2010.
[17] U. Fecker, M. Barkowsky, and A. Kaup, "Improving the prediction efficiency for multi-view video coding using histogram matching," Picture Coding Symposium (PCS), Apr. 2006.
[18] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," IEEE, Special Issue on Advances in Video Coding and Delivery. vol. 93, pp. 99–110, Jan. 2005.
[19] F. Isgro, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," IEEE Trans. Circuits Syst. Video Technol., vol.14, no. 3, pp. 288–303, Mar. 2004.
[20] A. Vetro, T. Wiegand and G. J. Sullivian, ''Overview of the stereo and and multiview video coding extension of the H.264/MPEG 4 AVC standard,'' Proceeding of the IEEE, vol. 99, no. 4, pp. 626-642, April 2011.

BIOGRAPHIES

	Shivam S. Agrawal joined the faculty of Electronics and Telecommunication at PLIT in 2016. He received his M.E. in 2016 from the Signal Processing at Smt. Kashibai Navale College of Engineering, Pune, also received undergraduate degrees in Electronics and Telecommunication at PLITMS, Buldana. He is currently working on Depth Video coding for 3D Video.
	Amol S. Bendmali joined the faculty of Electronics and Telecommunication at PLIT in 2015. He received his M.E. in 2014 from the Communication Networks at Sinhgad college of Engineering, Pune, where he also received undergraduate degrees in Electronics and Telecommunication. His recent work, besides Energy Efficient Communication, has focused on the: Electromagnetism. This has ranged from general research in semi analytical and numerical methods for Electromagnetism, to the analysis of integrated optical devices, to the analysis of optical fibers that guide light within an air core to circumvent limits of solid materials.