Reliable Stereoscopic Video Streaming Considering Important Objects of the Scene

Ehsan Rahimi and Chris Joslin

Department of Systems and Computer Engineering, Carleton University, 1125 Colonel By Dr., Ottawa, ON, Canada

Keywords: Stereoscopic Video, 3D/Multiview Video, Depth Map and Color Image, Multiple Description Coding, Error Prone Environment, Region of Interest, Pixel Variation, Coefficient of Variation.

Abstract: In this paper, we introduce a new reliable method of stereoscopic Video Streaming based on multiple description coding strategy. The proposed multiple description coding generates 3D video descriptions considering interesting objects contained in its scene. To be able to find interesting objects in the scene, we use two metrics from the second order statistics of the depth map image in a block-wise manner. Having detected the objects, the proposed multiple description coding algorithm generates the 3D video descriptions for the color video using a non-identical decimation method with respect to the identified objects. The objective test results verify the fact that the proposed method provides an improved performance than that provided by the polyphase subsampling multiple description coding and our previous work using pixel variation.

1 INTRODUCTION

Errors exist in the received video due to unreliable communication is one of the common problems that happens in both wired or wireless networks. In the wired networks, error can occur due to packet loss, corruption, congestion and large packet delay whereas in the wireless networks unreliable communication can stem from temperature noise and interference that exist in the physical environment. Also, when dealing with immersive videos, the increase of the data traffic load will consequently produce data congestion. Therefore, the serious packet failure problem needs to be addressed since such errors on the delivered video diminishes the viewing quality experience(Kazemi, 2012; Liu et al., 2015; Tillo and Olmo, 2007; Y. Yapc and Urhan, 2008; Ates et al., 2008; Wang and Liang, 2007; Wei et al., 2012). To avoid such errors, an error resilient method of data transmission is required used by the encoder.

Generally, there are usually three methods in the communication systems to avoid packet failure: Automatic Repeat reQuest (ARQ), Forward Error Correction (FEC), and Error Resilient Coding (ERC) (Kazemi, 2012). The first method, the ARQ approach requires a network with feedback capability and as a result it is not suited for real-time or broadcast applications. The second method, the FEC approach is designed to cope with a specific amount of noise error making it impractical for noise variances that exceed the threshold level. The third method, the ERC approach, is the approach of choice in this paper because of its resiliency against packet corruption or noise feature. This resiliency is achieved through redundancy bits added to the data stream. Among a number of ERC methods, the multiple description coding method is our method of choice due to its suitability for the channel with the large noise power. MDC avoids packet failure because it creates multiple complimentary and separately-decodable descriptions.

Using MDC, a video stream is partitioned into several separately decodable descriptions and transmitted to its respective receivers. In the receiver, there are two different types of decoder - the side decoder and central decoder. The receiver chooses one of the two decoders based on the availability of error free descriptions remaining. With the MDC method should an error occur in one description, it may be fixed when considering other error free descriptions.

This paper organizes as follows: a literature review regarding multiple description coding and how it can be applied on the stereoscopic video, is presented in Section 2. Then, the proposed method will be introduced in Section 3 and afterword, test results will be presented and discussed in Section 4. Finally, we have a review about our achievement in Section 5.

Rahimi, E. and Joslin, C.
Reliable Stereoscopic Video Streaming Considering Important Objects of the Scene.
DOI: 10.5220/0006616801350142
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 4: VISAPP*, pages 135-142
ISBN: 978-989-758-290-5

2 STATE OF THE ART

The MDC method is best recognized for its error robust property at the expense of compression ratio as it adds redundancies in its temporal, spatial or frequency domain. With the temporal MDC method, usually two descriptions are produced in order to avoid a drop in the coding efficiency. The drop in the coding efficiency is reflected when more than two descriptions are used because the distance between the assigned frames to each description is increasing resulting in the motion prediction being less effective (Liu et al., 2015; Chakareski et al., 2005). When the network is very noisy, a higher number of descriptions are required. Therefore the temporal MDC method is no longer a suitable technique. The frequency MDC method partitions Discrete Cosine Transform (DCT) coefficients between video descriptions. Because DCT transformation provides independent components, the descriptions will be less dependent. To maintain the correlation of the descriptions, extra transformation like Lapped Orthogonal Transformation (LOT) needs to be applied (Chung and Wang, 1999; Sun et al., 2009). Therefore the complexity of frequency MDC methods is higher than that of both the spatial and temporal MDC methods respectively. With the spatial MDC method, each video frame is partitioned into several lower resolution subimages using Polyphase SubSampling (PSS) algorithm (Shirani et al., 2001; Gallant et al., 2001; Kazemi, 2012). It is worth mentioning that with a simple spatial MDC method, there is no precise adjustment tool over the redundancy in order to control the side quality(Shirani et al., 2001; Gallant et al., 2001; Kazemi, 2012). This means that there is no control for the redundancy increase resulting in higher resistivity to compensate for the higher noise level.

To apply the MDC method for 3D videos, the depth image also needs to be partitioned into different descriptions. It is worth mentioning that the depth map image mainly contains depth information of the scene's objects. Because of the nature of the real objects, depth information of 3D scenes rarely contain high frequency content. Consequently, the depth map image can be effectively compressed effectively resulting in saved bandwidth and disk space (Fehn, 2004; Hewage, 2014). To improve compression, Karim et al. have shown that the downsampled version of the depth map image provides an adequate reconstruction of the 3D video in the receiver (Karim et al., 2008). They have experimented with the spatial MDC method for 3D videos using color plus depth map image representation. Karim et al. have carried out experimental tests with a scalable multiple description coding approach arriving at the same result. Therefore, it can be said that downsampling of the depth map image does not cause a considerable degradation in the quality of a reconstructed video. This is due to the fact that the depth map image includes low frequency contents or more precisely, the depth values of adjacent pixels are similar. Consequently, one can state that the neglected pixels during downsampling can be better predicted. Liu et al utilized the fact of having similar depth values of pixels for real objects and introduced a texture block partitioing algorithm in order to perform their MDC algorithm for wireless multi-path streaming (Liu et al., 2015).

However, multiple description coding has been investigated for 2D videos thoroughly. More investigation is required to apply MDC to 3D video specifically. For 2D videos, different MDC methods are classified according to the type of data which is divided into descriptions which include: temporal, spatial, frequency, or compressed. For example, with a temporal MDC method using two descriptions, one description can be odd frames and the other description even frames. With a spatial MDC algorithm each video frame is partitioned into several lower resolution subimages. With a frequency MDC method, the frequency components divide between descriptions. Each type of MDC method has its own advantages and disadvantages with regard to its particular application. The temporal MDC method is simple though unsuitable for an application involving a network with high packet failure due to its low capability in increasing data redundancy. With the higher complexity of frequency MDC method, the spatial MDC method can best accommodate a live HD video conference application over an error prone environment.

3 PROPOSED METHOD

This section describes the new proposed multiple description coding applicable for 3D videos considering ROI. In order to be able to recognize which part of the frame is more important or ROI map extraction, a metric needs to be defined. To this end, two metrics (PV and CV) are defined and the result for each metric will be compared at the end. For the first metric (PV), we calculated the average of the absolute variations for pixels' values found in the depth map image in a block wise manner:

$$PV_i = \frac{1}{N_i} \sum_{j=1}^{N_i} |D_j - \mu_i| \quad (1)$$

where μ_i is the average of depth values for block i and PV_i stands for the pixel variation of block i; D_j

is the depth value of pixel j in ith block and N_i is the total number of pixels in block i (i.e. $j = 1, 2, ..., N_i$). Generally, PV_i of block i is a Non-negative value that can vary from zero to infinity. Large PV shows that block i is probably related to several objects or edges and very small PV states that block i is likely related to the far distanced background or the planar objects for example, a wall. This is due to the fact that the depth information of an object contains low frequency contents, naturally.

For the second metric, we define a new metric (CV) as the ratio of Pixel Variation (PV) to the mean μ, also known as Coefficient of Variation (CoV):

$$CV_i = \frac{PV_i}{\mu_i}, \tag{2}$$

where CV_i is CoV for the block i within a depth map image. Like before, μ_i stands for the mean value of depth for the pixels in block i . PV_i has already been defined in Equation (1). Similar to PV, the CV is also a positive value. When CV of a block equals one then the depth values of that block have the same mean and standard deviation values. It can be argued that blocks with large CV values are probably related to several objects or edges while blocks with very small CV values are related to the background of the video frame. Consequently, they are not the interesting part of the frame that the ROI extraction algorithm is looking for.

Figure 1 shows an overview of the proposed encoder. As can be seen in this figure, the first step of the proposed encoder is to determine which part of the frame is more important. One important issue in this process is its requirement for a low complexity algorithm in order to realize the interesting objects in the frame. The ROI extraction algorithm proposed in this paper uses the characteristics of the depth map image and extracts the map of ROI using one of the metrics explained in the previous section. In this algorithm, the ROI range is defined as the distance between σ_{min} and σ_{max}. σ_{min} is the threshold which is used to separate the very far background objects from the interesting objects and σ_{max} is the limit used to detect edges of the interesting objects. Also N_{itr}^{Tot} is the total possible number of iterations that can be run by the hierarchical block division algorithm. The algorithm that identifies the objects is run in four major steps:

- **Step 1:** Create two empty lists (L_1&L_2), and assign the entire depth map image as one block to L_1. Then start the first iteration as explained in step 2.
- **Step 2:** Check if the algorithm reaches the limit of N_{itr}^{Tot} or if all blocks in L_1 are with PV or CV values smaller than σ_{max}^{PV} or σ_{max}^{CV}, respectively. If yes, go to step 4. If not, go to step 3. Clearly, in the first iteration there is only one block in L_1 and its metrics are with the strong probability greater than σ_{max}.
- **Step 3:** For every block in L_1 with the metric value greater than the threshold, divide the block into four equal sized blocks and assign them to L_2. Any block with metric value less than the threshold is assigned without change to L_2. After having checked all the blocks in L_1, L_1 is updated with L_2 and L_2 is cleared. Then return back to the step 2.
- **Step 4:** All blocks in L_1 with metric values less than σ_{min} are considered as region I. Blocks with metric values within the ROI range are considered as region II and remainders are region III .

In the hierarchical block division algorithm, a block is partitioned to smaller blocks by dividing the width and height of the block by a factor 2 in each iteration. It is worth mentioning that N_{itr}^{Tot} should be defined in order that the minimum block size be greater than a 2×1 or 1×2 pixels block size. This is due to the fact that both metrics used in this algorithm evaluate pixel variation where there is at least two pixels to measure the variation.

To have reliable video streaming, the proposed new spatial MDC algorithm exploits the Multiple Description Coding (MDC) strategy for 3D videos after ROI extraction algorithm. To this end, four descriptions are created using Poly phase SubSampling (PSS). PSS-MDC is the basic low complex method that can be used in the spatial domain to have a reliable transmission in the error prone environment. Although, the most important advantage of the PSS-MDC encoder is its simplicity, there is a capability lack in increasing the redundancy in order to avoid errors in the strong noisy environment. To fix this, the new spatial MDC algorithm enhances the pixel resolution for areas that are less predictable and also on objects of interest that are more important to focus on.

As can be seen in Figure 1, two different algorithms are applied on the color video and the depth map stream. For the depth map stream, the resolution of each description is enhanced according to its prediction difficulty. Since the metrics defined in this paper evaluate the variation between adjacent pixels, it can be said that pixels of the depth map frame are clustered into regions I to III according to their difficulty prediction levels. This means that the region I, which includes pixels with very low variations, remains without any change. Pixel resolution in the region II is enhanced to one second for each description. Since the region III contains pixels with large variati-

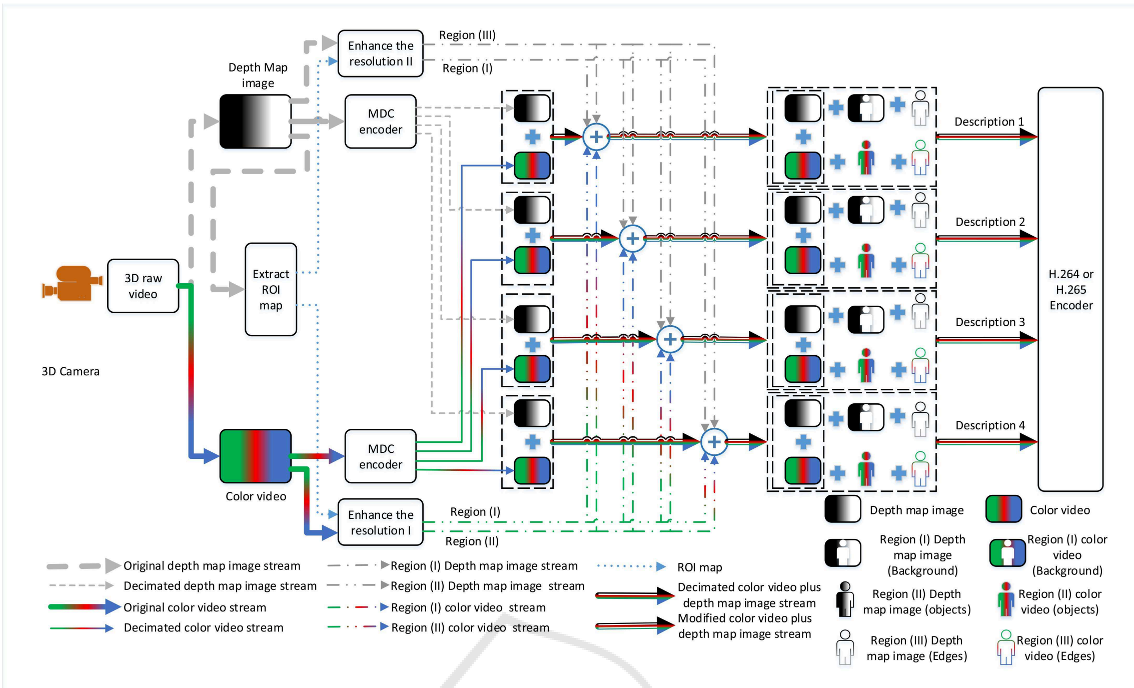


Figure 1: Block diagram of the proposed method.

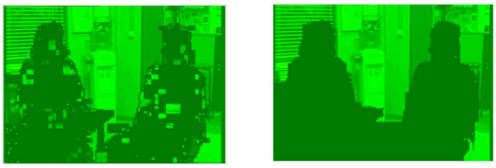
ons, it is likely that the prediction of a pixel (in case of missing) from adjacent pixels leads to error. As a result, this region's pixel resolution has increased to a fuller pixel resolution for each description.

Since the region's clustering algorithm is done using the depth map image rather than the color video frame, it cannot reflect the pixels' value variations for the color video frame. Therefore, the above mentioned argument is no longer applicable. One suggestion with regards to the color video is to apply the proposed ROI detection algorithm on the color video stream in order for it to extract ROI map based on the pixel variation found in the color video frame; but the drawback is its greater complexity due to a wide variety of colors inherently part of any scene naturally. As a result, the hierarchical block division algorithm needs more time to identify different regions in the frame. Another suggestion is to use the ROI map extracted from the depth map image to then focus on region II for the enhancement of pixel resolution in the color video frame rather than on region III which is performed within the depth map stream. Since the human eye is more sensitive to objects rather than of pixels, this suggestion introduces better performance with regards to the subjective assessment. Also, it can provide improvement with regards to the objective assessment since the recording of moving objects inherently part of the frame in the scene are now more focused. Because all video coding standards use Differential Pulse Code Modulation (DPCM) and proximate pixels' values of the objects in the color video frame, the increase of the resolution of those parts of a frame that include the ROI can be compensated by DPCM algorithm in point of compression ratio. Therefore, with regards to the color video stream, region II and III are enhanced to full and one second resolution, respectively. Region I remains with the same resolution as before (one fourth). This enhancement algorithm helps to perfectly recover the ROI in the instance of missing a description, although at the expense of increased redundancy.

4 SIMULATION RESULT AND DISCUSSION

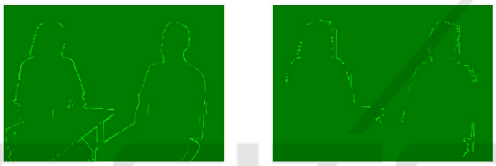
For the assessment of the proposed algorithm, this paper carried out some tests using two stereoscopic test video sequences with the format of DVD-Video PAL (720×576), called video "Interview" and "Orbi". Each video has 90 frames and the frame rate is 30 frames per second (fps). The chroma and depth subsampling format is 4: 2: 2: 4 (the last 4 stands for the resolution of the depth map image) or in other words the total frame resolution is 1440×576. The new algorithm is implemented using H.264/AVC reference software, JM 19.0 (Institut, 2015). To encode with JM software, I frames are repeated every 16 frames and only P frames are used between I frames.

(a) Original video frame.



(b) Region I (Ex. by *PV*). (c) Region I (Ex. by *CV*).

(d) Region II (Ex. by *PV*). (e) Region II (Ex. by *CV*).



(f) Region III (Ex. by *PV*). (g) RegionIII (Ex. by *CV*).

Figure 2: Comparison the performance of extracting different regions (I-III) for the first frame of video "Interview".

In the remainder of this paper, we will first investigate and compare the Performance and Complexity of the proposed algorithm using *PV* and *CV* . Then, we will assess the performance of the new proposed spatial MDC algorithm for streaming in the noisy environment. It is worth mentioning that to simulate an error prone environment, we have assumed that the decoder receives only one description among four descriptions generated in the encoder.

Figure 2 shows the identified regions I to III using *PV* and *CV* metrics. As can be seen, the identified region II is more accurately depicted with the *CV* metric rather than with the *PV* metric. The same scenario is also applicable for the region I. As can be seen in Figure 2d there are some important pixels that have not been detected as the region II (ROI). Also we have identified some missed pixels in region I (background) with *PV* as shown in Figure 2b. Such inaccuracy in realizing different regions with *PV* can be due to the fact that pixel values of different blocks are in dissimilar ranges. Therefore the pixel variation (*PV*) can not be an appropriate metric to be used when extracting for regions I and II. To fix this problem as argued before, it is necessary to normalize the pixel variation metric(*PV*). Indeed, the *CV* metric is the normalized version of pixel variation and works like a smoothing filter. Although using normalized pixel variation metric (*CV*) provides a considerable improvement in the extraction of regions I and II, such performance is not achieved when using the *CV* metric in detecting region III (which stands for the edges). As can be seen in Figure 2, the detected edges shown in Figure 2g is not as clear as the detected edges shown in Figure 2f. This can be due to the smoothing effect brought about by the normalization using the *CV* metric. As the blocks that contain edges are considered as blocks with high frequency contents, a high frequency filter like the pixel variation measurement (*PV*) is more beneficial for identifying the edges. Therefore, an optimum algorithm can extract the edges using metric (*PV*) an then detect the important objects using metric (*CV*).

Table 1 shows the average number of blocks for different metric values of *PV* and *CV*. As can be seen by this table, about 55% of the depth map image for video Interview and 40% of the depth map image for the video Orbi have *PV* values less than 1. On the other hand, for the video Interview more than one half and for video Orbi more than one third of the depth map image have very close depth values. This is the reason why the decimation of the depth map image does not affect its quality when it is reconstructed in the decoder. Table 1 also shows that about 95% of the depth map image for both test video sequences have *PV* values less than 3. The fact that about 95% of the depth map image have similar depth values result in no longer needing to send the depth map image with its original resolution, justifies why the non-identical decimation is more advantageous than the identical decimation sugested by Karim et al. in (Karim et al., 2008). On the other hand, only about 5% of the depth map image needs to be encoded with the original resolution. The 95% remainder can be decimated to save bandwidth or storage.

To investigate how robust the proposed MDC method is against error, we assumed that only one description is availble to decoder and all other three descriptions have been lost. In order to reconstruct the video, the decoder estimates the missed pixel value from the nearest available pixel value. Figure 3 and Figure 4 compare PSNR and SSIM measurements of the reconstructed color video for video Interview using the basic Poly phase SubSampling MDC method (PSS-MDC), our previous MDC method presented in (Rahimi and Joslin, 2017), and the new proposed spatial MDC algorithm with the help of *PV* and *CV* metrics. Figure 5 and Figure 6 also show

Table 1: Number of blocks with different metric values after hierarchical division algorithm.
(a) Video "Interview".

		Blocks' size								Percent of blocks with metric value in a specific range(%)
		6 (2×3)	24 (4×6)	96 (8×16)	384 (16×24)	1536 (32×48)	6144 (64×96)	24576 (128×192)	98304 (256×384)	
PV	≤ 1	662.78	371.67	172.54	75.80	22.28	17.82	0.68	0.00	55.68
	$1\sim 3$	1008.44	618.18	336.37	133.91	25.44	2.82	0.11	0.00	41.64
	$3\sim 10$	831.50	4.77	0.24	0.00	0.00	0.00	0.00	0.00	1.30
	≥ 10	898.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.37
CV	≤ 0.1	646.10	276.93	150.99	67.53	37.82	18.57	0.00	0.00	56.74
	$0.1\sim 0.2$	32.79	11.21	4.24	2.59	2.59	1.28	1.93	0.00	15.57
	$0.2\sim 0.3$	45.37	16.27	5.67	3.80	3.92	2.40	0.00	0.00	5.96
	$0.3\sim 0.4$	105.10	24.84	4.11	3.00	2.19	2.34	0.00	0.00	5.22
	$0.4\sim 0.5$	52.64	29.00	4.31	4.34	0.56	0.84	1.69	0.07	14.55
	≥ 0.5	1286.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.96

(b) Video "Orbi".

		Blocks' size								Percent of blocks with metric value in a specific range(%)
		6 (2×3)	24 (4×6)	96 (8×16)	384 (16×24)	1536 (32×48)	6144 (64×96)	24576 (128×192)	98304 (256×384)	
PV	≤ 1	542.72	295.40	172.56	69.13	34.42	4.69	0.80	0.00	39.37
	$1\sim 3$	1680.86	752.81	331.84	108.19	39.29	6.48	0.74	0.00	55.95
	$3\sim 10$	2276.38	8.09	0.47	0.00	0.00	0.00	0.00	0.00	3.53
	≥ 10	753.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.15
CV	≤ 0.1	614.22	244.82	118.68	49.28	35.30	6.58	0.81	0.00	39.28
	$0.1\sim 0.2$	59.43	28.10	11.99	8.81	5.51	1.64	0.10	0.00	6.76
	$0.2\sim 0.3$	79.78	28.24	9.36	5.84	4.41	2.56	0.48	0.40	19.80
	$0.3\sim 0.4$	134.88	35.41	10.13	4.74	4.08	1.38	0.22	0.61	21.54
	$0.4\sim 0.5$	90.23	41.59	9.84	3.82	2.31	0.80	0.00	0.30	10.66
	≥ 0.5	1285.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.96

the PSNR and SSIM assessments for the video Orbi. As can be seen in Figure 3, in the recreated video Interview about 1 dB improvement for the *PV* metric and 2 dB improvement for the *CV* metric can be achieved by the new proposed spatial MDC alorithm when compared to our previous work presented in (Rahimi and Joslin, 2017). Regarding video Orbi (see Figure 5), although a considerable improvement cannot be seen compared to our previous work, more than 2 dB improvement has been achieved by the new proposed spatial MDC algorithm in comparison with the PSS-MDC method. Regarding to the SSIM assessment, the proposed algorithm provides about 0.3 improvement for both test videos in high rate streaming compared to the PSS-MDC method. It should be mentioned that since the human eye is more sensitive to objects rather than that of pixels, a subjective assessment can better emphasize the improved performance brought forward by the proposed algorithm compared to the previous methods.

When it comes to the evaluation of the proposed algorithm for the reconstructed depth map image, it shows a better performance. As shown in Figure 7 and Figure 8 for the video Interview and in Figure 9 and Figure 10 for the video Orbi, the improvement of the proposed algorithm is considerably evident. This can be due to the fact that metrics *PV* and *CV* are calculated based on the depth map image and therefore blocks with larger values of metrics *PV* and *CV* can be considered as the least predictable blocks in the depth map image. Therefore, focusing on these pixels in each description results in a more accurate reconstruction in the decoder. In view of the PSNR assessment, about 8 dB for video Interview and more than 10 dB for video Orbi improvement have been achieved by the proposed algorithm. Such high performance of the proposed algorithm in view of the SSIM assessment is also more evident compared with the color video assessment. With regards to the SSIM assessment, the proposed algorithm outperforms by more than 0.02 compared to PSS-MDC method.

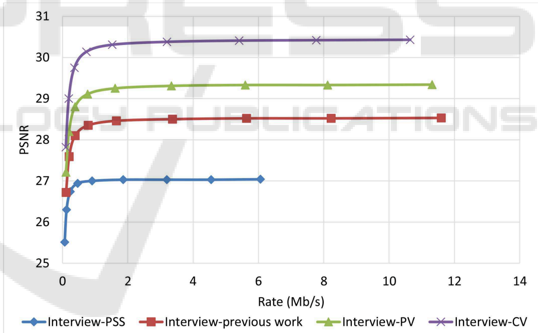


Figure 3: PSNR assessment of color image for video Interview.

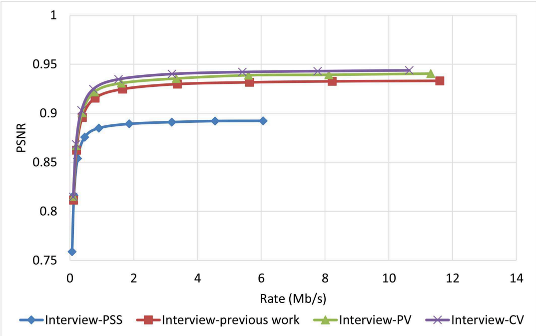


Figure 4: SSIM evaluation of color image for video Interview.

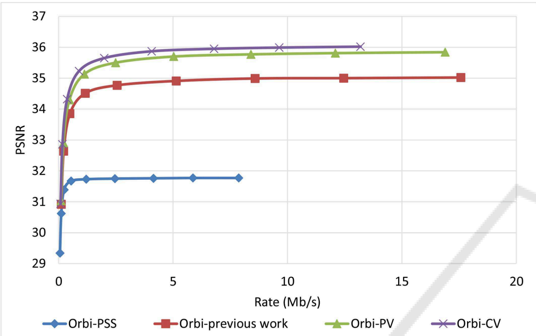


Figure 5: PSNR assessment of color image for video Orbi.

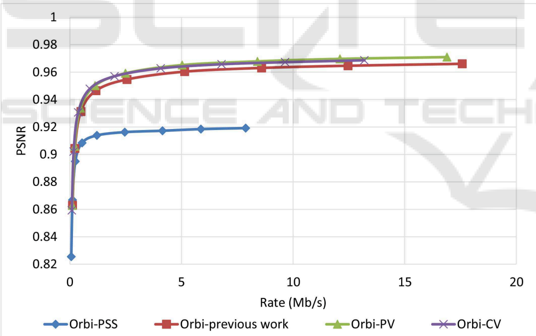


Figure 6: SSIM evaluation of color image for video Orbi.

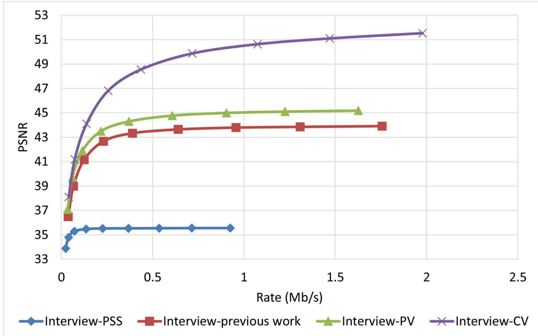


Figure 7: PSNR assessment of the depth map image for video Interview.

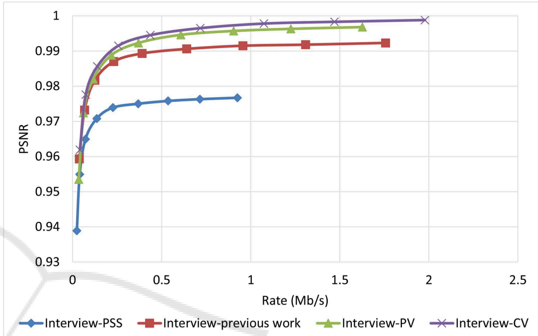


Figure 8: SSIM evaluation of the depth map image for video Interview.

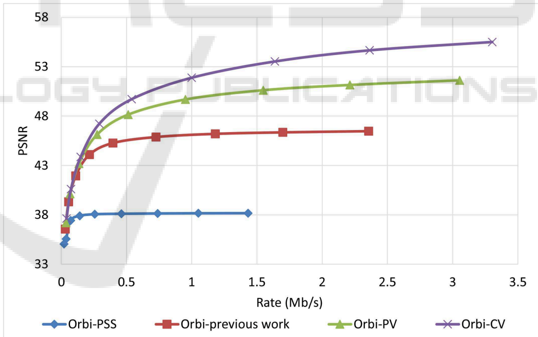


Figure 9: PSNR assessment of the depth map image for video Orbi.

5 CONCLUSION

Multimedia streaming is affected by packet failure in the network due to packet loss, packet corruption, and large packet delay. An appropriate solution against packet failure in the error prone environment can be multiple description coding (MDC). With MDC, one video description is partitioned into several separately decodable descriptions. In the instance of missing a description during transmission, the decoder is capable to estimate the lost description from other error free description(s). To improve the basic spatial partitioning and to be applicable to 3D videos, a non identical decimation algorithm for the stereoscopic videos has been provided in this paper. Our algorithm works based on existing objects in the scene and assigns more bandwidth to the region of interest.Since human eyes are more sensitive to the objects rather than that of pixels, the proposed algorithm can provide an improved performance compared to the PSS MDC method in view of subjective assessment. However, the objective assessment results confirm the improved performance achieved by the proposed spatial MDC algorithm. With regard to the depth map image, the proposed algorithm enhances the current basic decimation to a non identical decimation. As

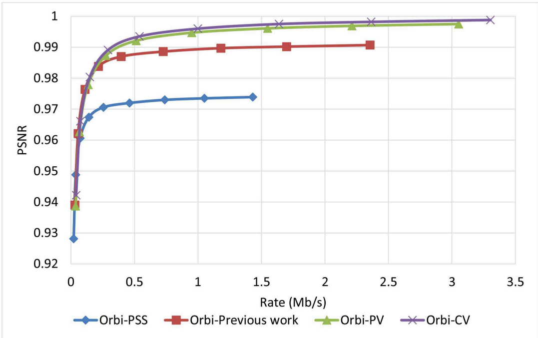


Figure 10: SSIM evaluation of the depth map image for video Orbi.

shown earlier, most parts of the depth map have similar depth values and therefore decimation in those parts can save bandwidth or storage without considerable quality degradation. However, for the parts of the frame with high pixels' value variation, it is recommended to keep the original resolution. Therefore, with the new algorithm those parts of the depth map image that have large variations is encoded with the original resolution.

ACKNOWLEDGEMENTS

The authors would like to acknowledge that this research was supported by NSERC Strategic Project Grant: Hi-Fit: High Fidelity Telepresence over Best-Effort Networks.

REFERENCES

Ates, C., Urgun, Y., Demir, B., Urhan, O., and Erturk, S. (2008). Polyphase downsampling based multiple description image coding using optimal filtering with flexible redundancy insertion. In *Signals and Electronic Systems, 2008. ICSES '08. International Conference on*, pages 193–196.

Chakareski, J., Han, S., and Girod, B. (2005). Layered coding vs. multiple descriptions for video streaming over multiple paths. *Multimedia Systems*, 10(4):275–285.

Chung, D.-M. and Wang, Y. (1999). Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms. *Circuits and Systems for Video Technology, IEEE Transactions on*, 9(6):895–908.

Fehn, C. (2004). Depth-image-based rendering (dibr), compression and transmission for a new approach on 3d-tv. *SPIE: Stereoscopic Displays and Virtual Reality Systems*, 5291:93– 104.

Gallant, M., Shirani, S., and Kossentini, F. (2001). Standard-compliant multiple description video coding. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 946–949 vol.1.

Hewage, C. (2014). *3D Video Processing and Transmission Fundamentals*. Chaminda Hewage and bookboon.com.

Institut, H.-H. (2015). H.264/avc reference software.

Karim, H., Hewage, C., Worrall, S., and Kondoz, A. (2008). Scalable multiple description video coding for stereoscopic 3d. *Consumer Electronics, IEEE Transactions on*, 54(2):745–752.

Kazemi, M. (2012). *Multiple description video coding based on base and enhancement layers of SVC and channel adaptive optimization*. PhD thesis, Sharif University of Technology, Tehran, Iran.

Liu, Z., Cheung, G., Chakareski, J., and Ji, Y. (2015). Multiple description coding and recovery of free viewpoint video for wireless multi-path streaming. *IEEE Journal of Selected Topics in Signal Processing*, 9(1):151–164.

Rahimi, E. and Joslin, C. (2017). 3d video multiple description coding considering region of interest. In *Accepted in 12th International Conference on Computer Vision Theory and Applications (VISAPP 2017)*.

Shirani, S., Gallant, M., and Kossentini, F. (2001). Multiple description image coding using pre- and post-processing. In *Information Technology: Coding and Computing, 2001. Proceedings. International Conference on*, pages 35–39.

Sun, G., Samarawickrama, U., Liang, J., Tian, C., Tu, C., and Tran, T. (2009). Multiple description coding with prediction compensation. *Image Processing, IEEE Transactions on*, 18(5):1037–1047.

Tillo, T. and Olmo, G. (2007). Data-dependent pre- and postprocessing multiple description coding of images. *Image Processing, IEEE Transactions on*, 16(5):1269–1280.

Wang, J. and Liang, J. (2007). H.264 intra frame coding and jpeg 2000-based predictive multiple description image coding. In *Communications, Computers and Signal Processing, 2007. PacRim 2007. IEEE Pacific Rim Conference on*, pages 569–572.

Wei, Z., Ma, K.-K., and Cai, C. (2012). Prediction-compensated polyphase multiple description image coding with adaptive redundancy control. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(3):465–478.

Y. Yapc, B. Demir, S. E. and Urhan, O. (2008). Downsampling based multiple description image coding using optimal filtering. *SPIE: journal of Electronic Imaging*, 17.