3D Video Spatiotemporal Multiple Description Coding Considering Region of Interest

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

Keywords: 3D Video, Stereoscopic Video, Dolor Image, Depth Map Image, Multiple Descriotion Coding, Temopral MDC, Spatial MDC, Hybrid MDC, Reliable Multimedia Streaming, Region of Interest.

Abstract: 3D video applications are being more favourable for observers as their requirements to receive and display 3D videos become more available recently. Therefore the demand for more processing power and bandwidth is increasing to stream and display 3D multimedia services in either the wired or wireless networks. Since channel failure has been always as an integral part of communication between a receiver and transmitters, a robust method of video streaming is always a hot topic for researchers. To make robustness against failure more stronger, it needs to increase redundancies however it destroys the coding and compressing efficiency. Therefore there is a trade-off problem between the coding efficiency and robustness of the stream. Among different methods of reliable video streaming, this paper introduces a new reliable 3D video streaming using hybrid multiple description coding. The proposed multiple description coding creates 3D video descriptions identifying interesting objects of the scene. To this end, a map for the region of interest is extracted from the depth map image first with a not complex algorithm compared to the available machine learning algorithm. Having realized region of interest, the proposed hybrid multiple description coding algorithm creates the descriptions for the color video using the advantages of both spatial and temporal multiple description coding methods; To this end, a non-identical decimation method concerning the identified objects assigns more bandwidth to those objects; second, background quality is improved with the temporal information. This way, first, the proposed method provides better visual performance as the human eye is more sensitive to objects than it is to pixels; second, the background is reconstructed with higher quality as it usually has a low movement and temporal information is a better choice to estimate the lost information. The objective test results verify the fact that the proposed method provides an improved performance than previous methods.

1 INTRODUCTION

3D displays have been favourable among customers since a very long time ago, from 1922 that "The Power of Love" was shown as the first 3D public display. This is because of enhancing objects' realization that a 3D video can produce; however, 3D video display was limited to the public displays due to either hardware or software limitations. To make the 3D display more predominant, video production companies and researchers have made their effort to mitigate such physical or processing limitations. Thanks to new hardware and software technological achievements, nowadays everyone has access to multimedia services everywhere using mobile devices. Therefore the demand for multimedia services is growing every day and the need for more bandwidth or a more efficient streaming method especially for 3D videos is highlighted.

As described by Smolic and Kimata in (Smolic and Kimata, 2003), a 3D video is "geometrically calibrated and temporally synchronized video data". In other words, a 3D display needs more resources such as bandwidth, processing power, and storage as the depth information needs to be streamed in addition to the colour image (Calagari et al., 2017). As the main core of the current 3D video encoding methods relies solely on 2D video coding standards, the lack of a specific 3D video coding standards is more sensible; Therefore, quite a few researches have been conducted to stream 3D videos more efficiently for many different applications and scenarios.

Generally, a video sequence needs to be encoded efficiently before streaming to save bandwidth as much as possible. In addition to the 2D video encoding algorithms, 3D video encoding algorithms need to be applied to remove inter-views correlation (As

Rahimi, E. and Joslin, C.
3D Video Spatiotemporal Multiple Description Coding Considering Region of Interest.
DOI: 10.5220/0009158304740481
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 4: VISAPP*, pages 474-481
ISBN: 978-989-758-402-2; ISSN: 2184-4321

2D video encoding algorithms removes only the spatial and temporal redundancies) (Chen et al., 2016). After compression, the video sequence streamed toward the receiver that can be from one or more different communication channels. Since various channels may deal with different reliability conditions, the video sequence may fail to be decoded and displayed successfully. Therefore, a robust streaming method is required to avoid such failure especially for the communication channel with a large variance of noise. As described in (Kazemi et al., 2014; Rahimi and Joslin, 2018), multiple description coding is a favourable method to stream video sequence reliably when the variance of noise is so large to use forward error coding (FEC) to correct the failure; however, there are several joint MDC and FEC algorithms that are not our method of choice because of its complexity and also assuming the variance of noise is too large to use FEC.

An MDC method partitions a video stream into several separately decodable descriptions and then they are streamed through the network. This way, first, if enough resource (for example bandwidth) is not available to receive all descriptions successfully, a subset of all descriptions can be received and decoded; so, It can be said that at least a lower quality version of the original video is available and will be displayed. Second, errors that happen in one or more descriptions can be fixed considering other error-free descriptions. These two advantages make MDC method as a powerful strategy to avoid packet failure in multimedia communication for either wired or wireless networks (Kazemi, 2012; Padmanabhan et al., 2003).

In contrast to the error resiliency aspect of the MDC method, the coding efficiency will be degraded since each description needs to be included some extra information as the header(Baccaglini et al., 2010). The cost of decreasing in compressing ratio is unavoidable as each description needs to be separately decodable. Although, that is not the only cost and the compressing ratio is decreasing more as the data in each description is not as dependent as it was in the original description. Therefore the differential pulse code modulation(DPCM) technique used by the encoder is not as efficient as it was before. To increase the coding efficiency, more correlated data is required to be assigned in one description, however, this weakens the estimation power of a missed description from other available descriptions. Therefore, there is an error resiliency-coding efficiency trade-off problem(Kazemi et al., 2014; Rahimi and Joslin, 2018).

The domain that is chosen for the partitioning video data determines the type of an MDC method, that can be spatial, temporal, or frequency type. It is worth mentioning among various MDC types, the temporal MDC is more common since it is very simple and provides better performance. However, when the noise variance is very large and more than two descriptions are required, its performance degraded dramatically and other MDC types are more favourable. In this paper, we proposed a hybrid MDC method that gains from both spatial and temporal MDC types' benefits.

This remain of this paper organizes as follows: a brief literature review and our motivation are presented in Section 2. Then, the proposed method will be introduced in Section 3 and afterward, test results will be presented and discussed in Section 4. Finally, we have a brief review of our achievement in Section 5.

2 STATE OF THE ART

Generally, temporal MDC type is more favourable since it is very simple to implement and also it provides a better performance against the network failure compared to the spatial or frequency MDC methods; however, the temporal MDC approach is more sensitive to the variance of the noise and also coding inefficiency is more severe for a temporal MDC type when the number of descriptions is more than two(Kazemi et al., 2014). Also, to have the best estimation power, a temporal MDC predicts the lost frame bidirectionally and therefore an MDC decoder needs to receive the later frame to predict the corrupted frame. This may cause it unsuited for applications that ”time” is very important and initially are designed to support live streaming (Rahimi and Joslin, 2018).

The second option is to use spacial MDC type as it is simple to implement and needs a decoder with lower complexity compared to the frequency MDC type; however, the quality of reconstruction of lost description is lower than the temporal MDC type. To improve its performance a nonidentical decimation algorithm has been proposed in (Rahimi and Joslin, 2018) which is designed initially for 3D videos and does not increase complexity significantly as required for a live streaming application. By this algorithm, interesting objects in the scene are detected, first, and then, they are assigned more bandwidth compared to other parts which are mainly the background. In addition to improvement measured by the objective assessment presented in that work, it can also provide much better performance in view of subjective assessment; because human eyes are more sensitive to the objects rather than that of pixels and it is more important to have a high-quality reconstruction of objects rather than each pixel of the frame; although, the background is reconstructed with a very low quality.

This paper proposed a hybrid solution to have a better performance for both interesting objects and also background in a very noisy environment compared to the current methods. In the proposed method background is streamed using a temporal MDC type while objects are streamed using the spatial MDC presented in (Rahimi and Joslin, 2018). On the other hand, since the background of a frame is the part that has lower movement therefore having adjacent frames can improve the performance of reconstruction for the background, significantly. It is worth mentioning due to the low movement of the background there is also no need to predict the lost frame bidirectionally and having the previous frames is sufficient enough to estimate the lost frame. The proposed method will be described in more detail in the next section.

3 PROPOSED METHOD

This section aims to describe the proposed MDC method which is useful for the unreliable communications with a large variance of noise. Our method of choice, briefly, combines the temporal MDC method with the spatial MDC method presented in (Rahimi and Joslin, 2018) with some modifications. As described in (Rahimi and Joslin, 2018), descriptions are created by a non-identical decimation algorithm and stream toward the receiver to increase the quality of reconstruction for the interesting objects. To this end, interesting objects are defined as objects that are on focus when the video is recorded and called Region of Interest (RoI). It is worth mentioning that to avoid an excessive increment in total required bandwidth, the method presented in (Rahimi and Joslin, 2018) decreases the bandwidth assignment for the background.

Figure 1 shows the block diagram of the proposed method. As can be seen in the Figure 1, in the first step the ROI map is extracted using the depth map image. This process will be explained in detail in Section 3.1; then, the descriptions are created as explain in Section 3.2. In the next step, descriptions are modified temporally and spatially as described in Section 3.3.

3.1 ROI Map Extraction

To be able to create the description, the encoder needs to extract RoI map. To this end, each frame of the depth map image is searched for RoI, first, and then descriptions are created with more emphasize on the RoI area. Therefore, RoI decoded with higher quality causing an increase in the quality of experience; because human eyes are more sensitive to objects rather than pixels. Additionally, pixels of an object have more similar values compared to pixels of different objects either in the colour image or the depth map image. Since pixels of an object are assigned to one description, each description is compressed more effectively.

As mentioned earlier, RoI of a frame defines as those parts of the frame that is usually on focus during recording. On the other hand, the depth information of an object includes low-frequency content data as an object can only locate in one place at a time. Therefore, the RoI can be extracted by filtering low-frequency parts of the frame which are not also located very far from the camera. To this end, we are using the hierarchical block division (HBD) algorithm presented in (Rahimi and Joslin, 2018) to find areas with similar depth values (which indicated as low-frequency data). The metric used in this paper to extract RoI map is *Coefficient of Variation(CoV)* which outperform *Variance* as shown by simulation results presented in (Rahimi and Joslin, 2018). It is worth mentioning that CoV is defined as the ratio of the standard deviation to the mean (Curto and Pinto, 2009) and calculated by the following equation for each block:

$$CoV^{B_k^l} = \frac{\sigma^{B_k^l}}{\mu^{B_k^l}}, \tag{1}$$

where B_k^l stands for k^{th} block in l^{th} iteration of HBD algorithm; k can vary from one to the total number of blocks in each iteration . For example in the first iteration, the total number of blocks is one. We will talk more about the range of l and k later. $\sigma^{B_k^l}$ and $\mu^{B_k^l}$ are also the standard deviation and the average of k^{th} block in l^{th} iteration of the depth map image , respectively, i.e.:

$$\sigma^{B_k^l} = \sqrt{\frac{1}{N^{B_k^l} M^{B_k^l}} \sum_{i=1}^{N^{B_k^l}} \sum_{j=1}^{M^{B_k^l}} (d_{ij}^{B_k^l} - \mu^{B_k^l})^2}, \tag{2}$$

and

$$\mu^{B_k^l} = \frac{1}{N^{B_k^l} M^{B_k^l}} \sum_{i=1}^{N^{B_k^l}} \sum_{j=1}^{M^{B_k^l}} d_{ij}^{B_k^l}, \tag{3}$$

where $N^{B_k^l}$ and $M^{B_k^l}$ are the number of columns and rows of B_k^l, respectively and $d_{ij}^{B_k^l}$ is the depth value of the pixel located at column i and row j of B_k^l.

Generally, CoV of a block is a positive value varies from zero to infinity. As discussed in (Mirahsan et al.,

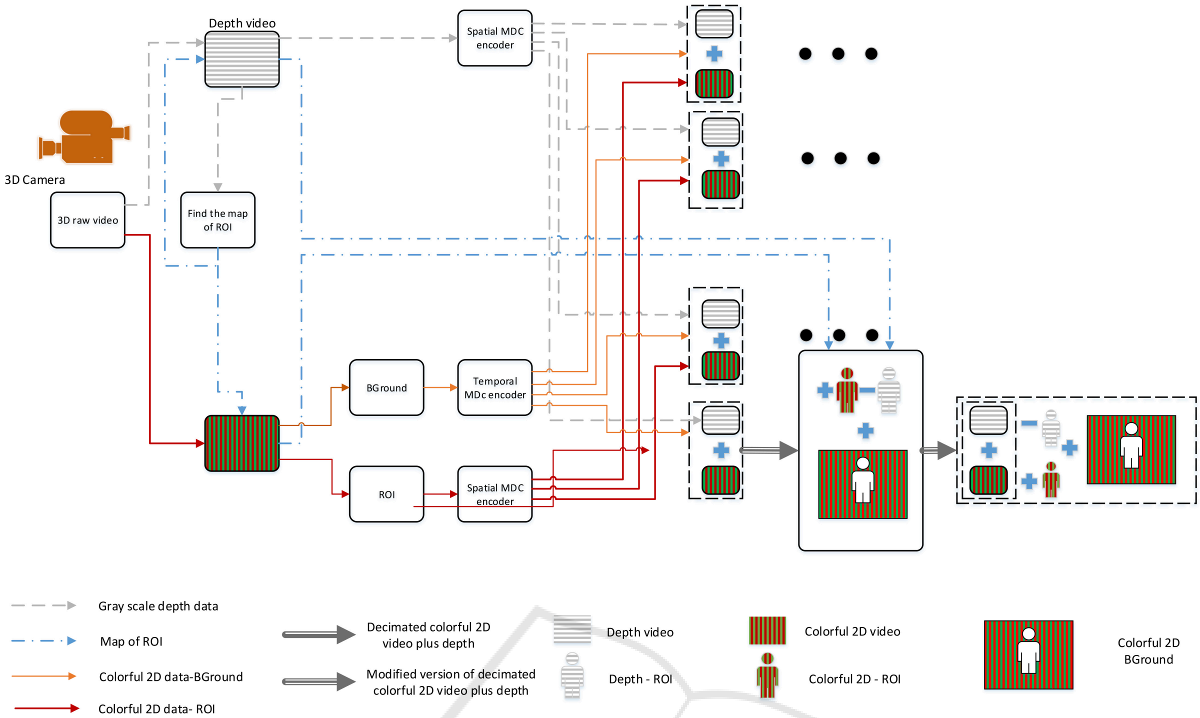


Figure 1: Block diagram of the proposed method.

2018) and (Mirahsan et al., 2014), *CoV* can be used as a measurement tool of the level of the heterogeneity or clustering for a random process. They showed that a realization of a process with *super-Poisson* characteristic results in *CoV* greater than one, with *sub-Poisson* characteristic results in *CoV* less than one, and with *Poisson* characteristic results in *CoV* equals one. Because the depth values of all pixels for an object are very similar, the *CoV* metric can be used as a benchmark metric to determine the level of clustering and consequently objects in the depth map image. It is worth mentioning that they are many neural network or learning algorithms to extract objects; however, having very complex algorithms make them unpractical for live stream application. Moreover, those algorithms need to be trained first to be able to recognize specific objects, while the proposed algorithm does not require training.

HBD algorithm separates the low-frequency parts of the frame by extracting out the blocks that have *CoV* of smaller than one in each iteration. To this end, the HBD algorithm considers each depth map image as one block initially and then decides to continue or stop the block division process. If *CoV* value is greater than one (and usually it is in the first iteration), then the block is portioned into four left/right-top/bottom equal size blocks. This means the total number of blocks increases from one in the first iteration to four in the second iteration. This process continues until all blocks have *CoV* values less than one or it is not possible to partition the block anymore (i.e. the block's size is 2×2). At the end of this process, each frame has several small, medium, and large size blocks with different *CoV* values.

As discussed in (Rahimi and Joslin, 2018), the HBD algorithm is stopped for about 5% to 9% of the entire depth map image of the test video sequences after the second iteration and also it is stooped for about one-third of the depth map image of both test videos after the third iteration. As argued in (Rahimi and Joslin, 2018), the results show that the HBD algorithm does not give rise to a high load of calculation.

As explained earlier, now the blocks with *CoV* less than one are considered as RoI and their top-left pixel's position plus their size are recorded as the depth map image. Having known the RoI map, encoder begins the process of creating descriptions which explained in the next subsection.

3.2 Description Creation

To create the description, each frame is partitioned spatially into for parts using *poly-phase sub-sampling* (PSS). To this end, from every adjacent 2×2 pixels, each description includes one pixel. The PSS process is shown in Figure 2. As shown in Figure 2, descriptions one, two, three and four include top-left, top-right, bottom-left, and bottom-right pixel of every ad-

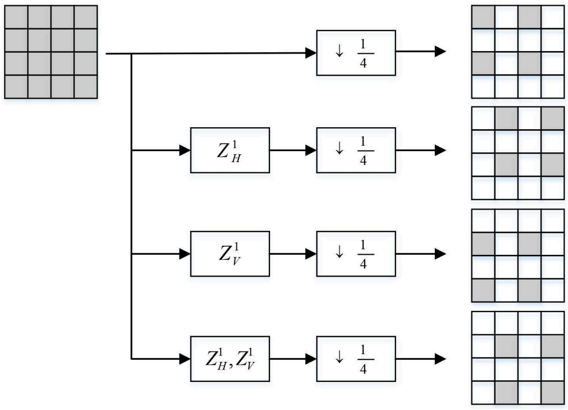


Figure 2: Polyphase SubSampling process: Z_H^1 and Z_V^1 are horizontal and vertical shift, respectively.

jacent 2×2 pixels, respectively. It should be noted that this process is done separately for the colour image and the depth map image.

3.3 Description Modification

At this step, each description is modified in a way so that RoI is assigned more bandwidth when it streamed. It is worth mentioning that such modification is done in both temporal and spatial domains described in Section 3.3.1 and Section 3.3.2, respectively. In more detail, the spatial modification enhances only RoI part resolution and the temporal modification augments only not Roi part in which mainly includes background.

It is worth mentioning, for the depth map image the proposed method uses only the spatial modification process. This is because the quality of reconstructed depth map image frames using only spatial modification is high enough and there is no need to increase the complicity of the proposed method anymore. As shown in (Rahimi. and Joslin., 2018; Rahimi and Joslin, 2017), only spatial modification improves PSNR assessment of the reconstructed video about 8-10 dB for using test video sequences.

3.3.1 Spatial Modification

The main goal of this part is to increase the redundancy of each description to make it robust enough to avoid failure in the strong noisy environment. An increase in redundancy needs to be proportional to the importantness of the block. As discussed earlier, it has been assumed the most important part of a frame is that part that is in focus when the video is recorded.

Now, when the HBD algorithm finishes if a bock is part of RoI, its resolution is increased from one-fourth to full resolution otherwise its resolution is kept as it was before. Such modification affects the video in two ways:

- First, since human eyes are more sensitive to objects rather than pixel it can provide a better quality of experience as such main objects are reconstructed with better quality.
- Second, since pixels values of an object, are usually similar and due to the use of *differential pulse code modulation* (DPCM) in the encoder, increasing the resolution of an object does not increase the volume of encoded description significantly.

It should be noted that such modification is only applied to color image frames. As the RoI part includes pixels with very low variations of the depth map image, a reverse spatial modification needs to be applied on depth map image frames. This means the resolution of the non-RoI part is increased from one-fourth to full resolution and the resolution of the RoI part remains unchanged.

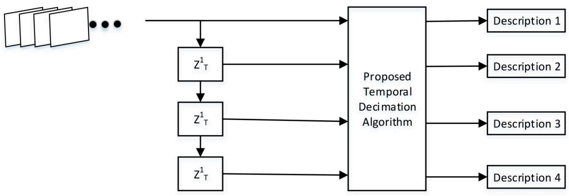
3.3.2 Temporal Modification

With the spatial modification process, only main objects can be reconstructed with high quality and background objects are reconstructed with low quality. It is worth mentioning that background is the part of a frame that usually has low movement and therefore a missed pixel can be predicted more accurately from the pixel located at the same position from the adjacent frames instead of looking surrounding pixels at the same frame. Figure 3 show how the temporal modification process augments each description information. As shown in Figure 3a for every four frames each description increase the resolution of not RoI part to the full resolution. For example, in description one, only frames $1, 5, \ldots$ have full resolution information of the background. To avoid a huge increase in the volume of data in each description, the first and third frames after the full resolution frame are completely dropped from each description and they are predicted from the adjacent frames. Because of the low movement of background, the missed frame can be reconstructed with higher quality than it was before (as presented in (Rahimi and Joslin, 2018)). Again, such temporal modification only applies to the color image.

4 SIMULATION RESULT

For the evaluation of the proposed algorithm, this paper carried out several tests using two stereoscopic video sequences with the format of DVD-Video PAL(720×576), called video "Interview" and

(a) Temporal decimation algorithm.



(b) Temporal multiple description coding.

Figure 3: Temporal modification process: Figure 3a shows the temporal decimation process (green: full resolution, yellow: as was before, red: dropped) and Figure 3b describes how descriptions ar created.

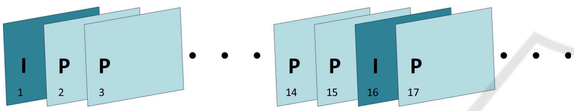


Figure 4: Reference frame stricture of the H.264/AVC encoder.

”Orbi”. Both test video sequences have chroma and depth subsampling format as 4: 2: 2: 4 (the last 4 shows the resolution of the depth map image). Each sequence includes 90 frames and the frame rate is 30 frames per second (fps). The new algorithm is implemented using H.264/AVC reference software, JM 19.0 (Heinrich-Hertz-Institut, 2015). To encode with JM software, I frames are repeated every 16 frames and only P frames are used between I frames as shown in figure 4.

It is worth mentioning that for both test sequences, the width of the depth map frame ($720 = 2^4 \times 3^2 \times 5$) is not divisible after the fourth iteration. To have a fine resolution for each block it is required to continue the HBD algorithm for more iterations. Therefore, the depth map frame resolution needs to modified from 576×720 to 512×768. It means that the last 64 rows of pixels are cropped and newly added 48 columns are zero-padded. This way the minimum possible block size can be 2×3 which is achieved after the eighth iteration. Also, it should be noted that to simulate an error-prone environment, it is assumed three descriptions are lost during communication and the decoder receives only one of four descriptions generated in the encoder.

First, we apply the bi-directional temporal MDC method and basic spatial MDC method applied to both video sequences to compare the performance of temporal and spatial MDC. Figure 5 shows the PSNR assessment of both MDC types with two or four descriptions. In this test, the colour image is partitioned spatially and temporally into two, four descriptions and streamed toward the receiver. In the decoder, it is assumed that only one description is available and others are missed. Then the missed information is estimated from the available description. As shown in Figure 5, temporal MDC performs much better than spatial MDC in point of PSNR assessment, however, spatial MDC is more robust to the noise variation compared to the temporal MDC. As can be seen in Figure 5 the slop of the graphs related to the spatial MDC for the large rate is approximately zero while the slope for the temporal MDC is not zero. This means that temporal MDC is more sensitive to the noise compared to spatial MDC. Also, It should be noted that these results are the average PSNR assessment of all 90 frames. In fact, the successfully received frames are decoded with a PSNR assessment around 54 dB for the large rate while the missed frames are decoded with PSNR about 37 dB. So there is huge between the frame that received successfully and the one estimated. In comparison, for the spatial MDC, all frames are decoded with PSNR assessment approximately 38 dB. It is worth mentioning that the test video sequence selected for this test has very low movement and frames are very dependent; therefore the temporal MDC provides much better results. Moreover, in this test, the temporal MDC is using both previous and next adjacent frames to reconstruct the missed frame that is not favourable for the applications explained earlier. Therefore, the proposed method is using the spatial MDC and also tries to improve the performance by increasing the ROI bandwidth and adding temporal information for the not RoI parts.

Figure 6 and Figure 7 show the PSNR assessment of the color image for video ”Interview” and ”Orbi”, respectively. The graphs compare the performance of the proposed method and previous methods. As can be seen in Figure 6, the quality of the reconstructed video of the test video sequence ”interview” is improved by about 6 to 7 dB. First of all, this huge gap is due to the very low movement of the background that we have in this test video sequence; therefore adding temporal information improves the PSNR assessment significantly. As argued before, the slop of PSNR assessment for the proposed method is smaller than the slope of the temporal MDC and greater than the slope of the spatial MDC presented in Figure 5; so it is less sensitive to the noise variation compared to a pure temporal MDC type. More importantly, the proposed method provides better performance compared to the spatial MDC type as we were looking for. In

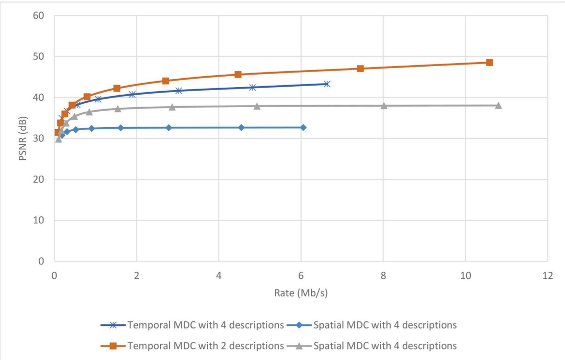


Figure 5: Temporal MDC vs. Spatial MDC performance comparison for the video ”interview”.

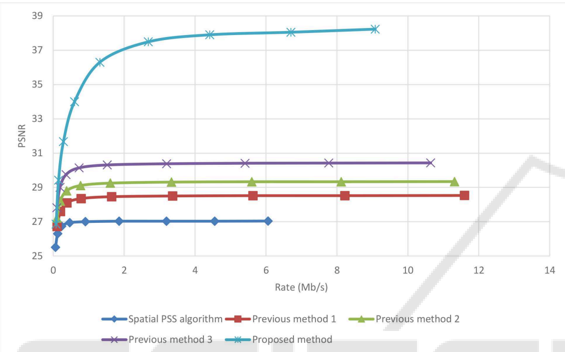


Figure 6: PSNR assessment of the color image for video ”interview”.

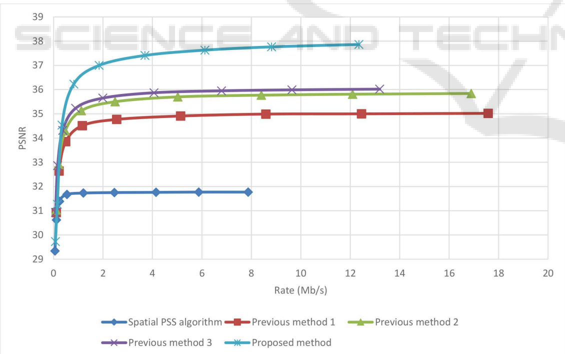


Figure 7: PSNR assessment of the color image for video ”Orbi”.

Figure 7, we observe similar results for the test video ”Orbi”, however, the improvement is not as large as the improvement achieved for the test video ”Interview”. As shown in Figure 7, the proposed method outperforms about 2 *dB* compared to previous methods.

As explained in Section 3, the proposed method does not modify the algorithm presented in (Rahimi and Joslin, 2018) for the depth map image because its performance is high enough and there is no need to increase the complexity of the proposed method. As

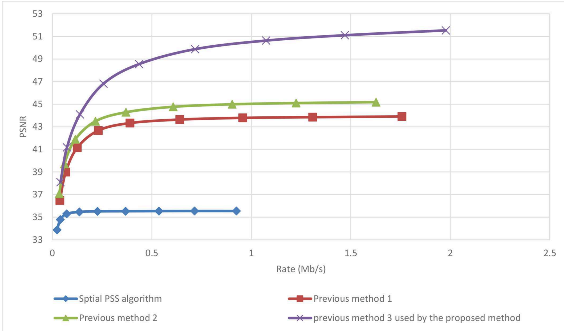


Figure 8: PSNR assessment of the depth map image for video ”interview”.

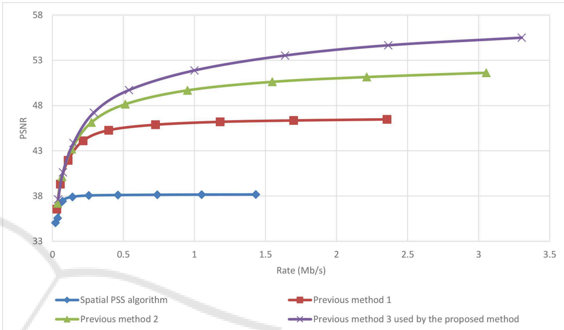


Figure 9: PSNR assessment of the depth map image for video ”Orbi”.

can be seen in Figure 8 and Figure 9, the PSNR assessment of previous algorithm is more than 45 *dB*. It is worth mentioning that in practice PSNR assessment greater than 40 *dB* is high enough to have user satisfaction. Therefore, we just reproduce the result presented in (Rahimi and Joslin, 2018).

5 CONCLUSIONS

To have a reliable multimedia communication, it is required to add redundancies before streaming the multimedia service through the network. As argued, MDC is one of the most favourable methods that avoid multimedia services’ delivery failure and is our method of choice in this paper. Current MDC methods usually create descriptions without considering the content of the video. This paper proposes a hybrid spatial and temporal MDC methods considering the content of the video. To this end, we first create spatial MDC descriptions and then the spatial information of ROI parts is augmented. Applying the spatial modification on the simple spatial PSS, the RoI part is decoded with higher quality and therefore the quality of experience of users is increased. It is worth mentioning that the proposed method can also provide improved subjective assessment since human eyes are more sensitive to objects rather than pixels; however, the non-RoI area, which is usually considered as the background, is decoded with a lower quality. To fix that, we also use the temporal MDC method in which the lost information of the non-RoI area (which is usually background) is estimated from available temporal information (previous frames). Since the background region is part of the frame that usually has very low movement compared to adjacent frames, the temporal domain is a perfect choice to be used for the estimation of lost data. The simulation results presented in Section 4 verify our argument.

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

Baccaglini, E., Tillo, T., and Olmo, G. (2010). A comparison between ulp and mdc with many descriptions for image transmission. *IEEE Signal Processing Letters*, 17(1):75–78.

Calagari, K., Elgharib, M. A., Shirmohammadi, S., and Hefeeda, M. (2017). Sports vr content generation from regular camera feeds. In Liu, Q., Lienhart, R., Wang, H., Chen, S.-W. K.-T., Boll, S., Chen, Y.-P. P., Friedland, G., Li, J., and Yan, S., editors, *ACM Multimedia*, pages 699–707. ACM.

Chen, Y., Zhao, X., Zhang, L., and Kang, J. (2016). Multiview and 3d video compression using neighboring block based disparity vectors. *IEEE Transactions on Multimedia*, 18(4):576–589.

Curto, J. D. and Pinto, J. C. (2009). The coefficient of variation asymptotic distribution in the case of non-iid random variables. *Journal of Applied Statistics*, 36(1):21–32.

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

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.

Kazemi, M., Shirmohammadi, S., and Sadeghi, K. H. (2014). A review of multiple description coding techniques for error-resilient video delivery. *Multimedia Systems*, 20(3):283–309.

Mirahsan, M., Senarath, G., Farmanbar, H., Dao, N. D., and Yanikomeroglu, H. (2018). Admission control of wireless virtual networks in hethetn ets. *IEEE Transactions on Vehicular Technology*, 67(5):4565–4576.

Mirahsan, M., Wang, Z., Schoenen, R., Yanikomeroglu, H., and St-Hilaire, M. (2014). Unified and non-parameterized statistical modeling of temporal and spatial traffic heterogeneity in wireless cellular networks. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 55–60.

Padmanabhan, V., Wang, H., and Chou, P. (2003). Resilient peer-to-peer streaming. In *Network Protocols, 2003. Proceedings. 11th IEEE International Conference on*, pages 16–27.

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)*.

Rahimi, E. and Joslin, C. (2018). Reliable 3d video streaming considering region of interest. *EURASIP Journal on Image and Video Processing*, 2018(1):43.

Rahimi., E. and Joslin., C. (2018). Reliable stereoscopic video streaming considering important objects of the scene. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4 VISAPP: VISAPP,*, pages 135–142. INSTICC, SciTePress.

Smolic, A. and Kimata, H. (2003). Report on status of 3dav exploration. Technical Report W5877, ISO/IEC JTC1/SC29/WG11, Norway.