International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 04 – Issue 07 || July 2019 || PP. 76-85

Scalable 3D video with Reduced Resolution Depth Compression

U. A. Sadiq
Centre for Satellite Technology Development
National Space Research and Development Agency (NASRDA), Abuja, Nigeria

A. H. Sadka
Department of Electronic & Computer Engineering
Brunel University
Uxbridge, London, UK

Abstract: Variable-quantization algorithms proved efficient in reducing the fluctuations of the output bit rate of a H.264-based Scalable Video Encoder. However, in applications where the bit rate budget is extremely low, the traditional rate control algorithms would fail to produce acceptable results since the quantization step size cannot be increased beyond an upper bound (i.e 31). In this case, the variable Qp techniques could be used in conjunction with the reduced resolution scheme to achieve a more efficient bit rate regulation process. The reduced resolution consists of down-sampling each image macroblock in the prediction error before it is encoded and up-sampling the reduced-resolution reconstructed block at the decoder in order to produce the motion compensated picture. Each down-sampled 8x8 luminance image block is transformed using 8x8 DCT.This proposed method reduces the overall bit rate and consequently improves rate distortion for 3D video at low bit rates in error free channels and improves 3D scalability performance for 3D transmission under high error channels.

I. INTRODUCTION

The reduced resolution depth coding is designated for use under very tight bit rate budget considerations. Since the reduced resolution depth frame results in a lower number of coefficients, the output bit rate per frame is reduced, while keeping a constant quantization parameter and a constant frame rate [1]. If this coding is to operate on an Macroblock basis, then an additional bit per macroblock is required to indicate its use. This incurs an extra bit in the bit stream for every coded macroblock, giving a maximum of 99 bits per frame for a QCIF resolution sequence [1][2]. For this reason, the down-sampling mode is usually decided on a frame-by-frame basis, and therefore a single extra bit per frame is then required in the frame header to indicate its use.

For extremely low bit rate budget, both variable-Qp and reduced resolution depth algorithms operate simultaneously to guarantee a smooth output bit rate without compromising the quality of the decoded video. The quantizer is selected with a bit rate prediction algorithm in[3] could be used to estimate the quantiser step size of a video frame. If the estimated bit rate exceeds the target bit rate of a frame after applying the feed-forward rate controller, then the reduced resolution controller is switched ON and a new quantization step size is estimated. The switching strategy is done on a frame by frame basis to reduce the complexity that is due to unnecessarily using the reduced resolution mode when the estimated bit rate falls below the target one. As describe before, the feed forward method tries to assign a fixed number of bits to each frame by selecting a quantization step size Qp that can achieve the target bit rate[3][4][5].

This paper investigates the application of reduced resolution depth data compression in SVC for 2D video plus depth stereoscopic video. The depth video frames are spatially down-sampled before SVC encoding and up-sampled after decoding. Section 2 provides a brief literature review on down-sampling and up-sampling (DSUS). Section 3 describes the proposed scalable 3D video with reduced resolution. Section 4 provides evaluation criteria for depth coding. Experimental results showing R-D performance under error free and error-prone conditions are depicted in section 5. The paper is finally concluded in section 6.

.

II. LITERATURE REVIEW

A. interview prediction with down-sampled reference images

Here, the encoder and decoder process the difference resolution videos[6]. We assumed that the spatial resolution of B views is smaller than I or P views, depicting in figure 1 below. For the prediction structure in the figure 2, 1 and P views are encoded in the same way as the JMVM[7][8].The proposed coding method is as follows, I and P are encoded in the same way as the JMVM and stored decoded images. Then when a picture in B sample them for use of inter-view prediction, and then it encodes the views the encoder is illustrated in the figure 2. The down-sampling process is indicated below.

Num_downsampled_views_minus_1 identifies the total number of views that require down-sampled reference pictures for inter-view prediction. The value of the number_of_downsampled_view_minus_1 shall be in the range of 0 to 1023. downsampled_view_id[1] specifies the view_id of the view that require down-sampled reference pictures. Post-processing especially at the encoder could control the up-sampling process as a post-processing. In the up-sampling process, not only decoded images of the current view, but also decoded images of other views could be used. The parameters for such process may be indicated as a message. Basically, down-sampling process increase complexity so also up-sampling process. The final goal of up-sampling is to obtain the same resolution for a lower bit rate

Table 1: Syntax for down-sampled reference images

seq_parameter_set_mvc_extension( ) {	**C**	**Descriptor**
num_views_minus_1		ue(v)
num_downsampled_views_minus_1		ue(v)
for(i = 0; i <= num_views_minus_1; i++)		
view_id[i]		ue(v)
for(i = 0; i <= num_views_minus_1; i++) {		
num_anchor_refs_l0[i]		ue(v)
for(j = 0; j < num_anchor_refs_l0[i]; j++)		
anchor_ref_l0[i][j]		ue(v)
num_anchor_refs_l1[i]		ue(v)
for(j = 0; j < num_anchor_refs_l1[i]; j++)		
anchor_ref_l1[i][j]		ue(v)
for(i = 0; i <= num_views_minus_1; i++) {		
num_non_anchor_refs_l0[i]		ue(v)
for(j = 0; j < num_non_anchor_refs_l0[i]; j++)		
non_anchor_ref_l0[i][j]		ue(v)
num_non_anchor_refs_l1[i]		ue(v)
for(j = 0; j < num_non_anchor_refs_l1[i]; j++)		
non_anchor_ref_l1[i][j]		ue(v)
for(i = 0; i <= num_downsampled_views_minus_1; i++)		
downsampled_view_id[i]		ue(v)

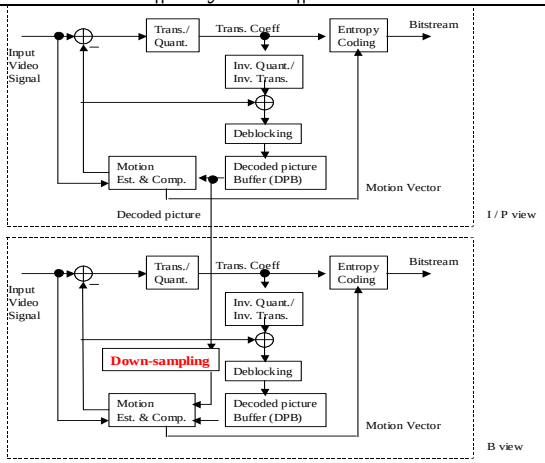


Figure 1: Encoder components when down-sampling is conducted in encoding B view

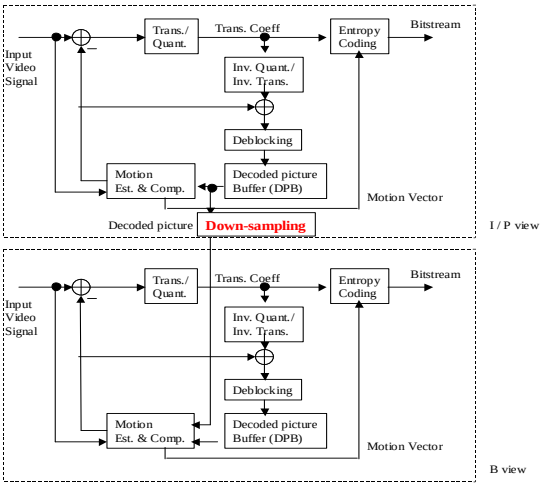


Figure 2: Encoder components when down-sampling is conducted in encoding I or P view

B. depth down-sampling/up-sampling (DSUS)

Encoding a reduced resolution depth can reduce the bitrate substantially, but the loss of resolution also degrades the quality of the depth map, especially in high frequency regions such as the object boundary; the resulting image rendering artifacts could be very visible and disturbing. Conventional down/up-samplers use a low pass filter and an interpolation filter to reduce the quality degradation[9][10]. However, since the depth video and image rendering results are quite sensitive to variations in space and time, especially on the object boundary, these traditional techniques are not sufficient.

Considering the above, we design a new down/up sampler for depth. In down sampling a 2D image, a representative value among the values in a certain window must be selected; we choose a median value.

$$imgdown(x, y) = median(Wsxs) \qquad (1)$$

Where Wsxs represents a sxs block and s is a scaling factor for down-sampling. That is, the down sampling reduces the image size by selecting the median value for each Wsxs.

The up-sampling process consists of the following steps: 1) image up-scaling, 2) 2D median filtering, 3) proposed depth reconstruction filtering. The up-scaling is an inverse process of the image down-sampling. We use same scaling factor in down sampling process

$$imgup(x, y) = imgdown(\left|\frac{x}{s}\right| \left|\frac{y}{s}\right|) \qquad (2)$$

After the up-scaling, we apply 2D median filter to smooth the blocking artifact caused by image down-sampling. Finally, the proposed depth reconstruction filtering is applied to reconstruct the object boundary distorted by down sampling.

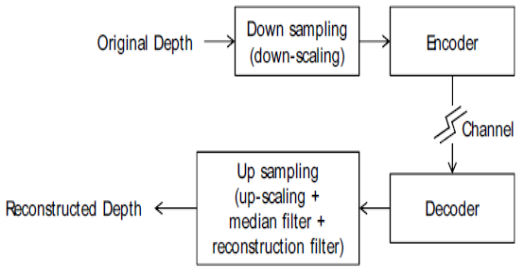


Figure 3: Flow diagram of depth down/up sampling

III. PROPOSED SCALABLE 3D VIDEO WITH REDUCED RESOLUTION DEPTH

The proposed method investigates the compression of depth information at reduced resolution for low bit rates. The depth video frames are down-sampled before encoding and up-sampled after decoding. This application of reduced resolution has been investigated in [11] with the sole aim of improving performance. The application of reduced resolution for depth image sequences is investigated in [12]. The reduced resolution is used in Mobile applications with the aim of reducing the bandwidth and improving the performance.

A simple way to down-sample an image by a factor of two is by using sub-sampling. If $f(i,j)$ is the pixel value of an image at location (i,j), then the down-sampled image is:

$$fd\left(\frac{i}{2},\frac{j}{2}\right) = f(i,j), \qquad (3)$$

$$for\ i = 0,2,4,\dots XSIZE, j = 0,2,4,\dots YSIZE$$

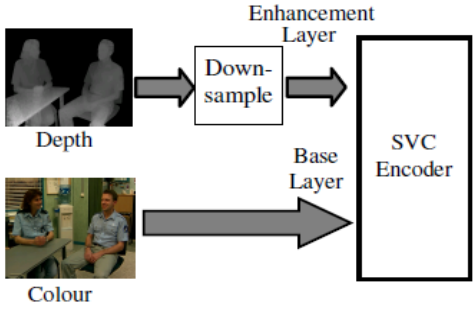
$XSIZE$ is the vertical size and $YSIZE$ is the horizontal size of the image to be down-sampled. It should be noted that three neighboring pixels to $f(i,j)$, which are $f(i,j+1), f(i+1,j)$ and $f(i+1,j+1)$, can also be used. In this paper, the pixel $f(i,j)$ plus the three neighborhood pixel values are averaged as below to produce an image down-sampled by a factor of two

$$fd\left(\frac{i}{2},\frac{i}{2}\right) = [f(i,j) + f(i,j+1) + f((i+1,j) + f(i+1,j+1)]/4 \quad (4)$$

$$for\ i = 0,2,4 \dots XSIZE - 1, j = 0,2,4 \dots YSIZE - 1$$

As an example, equation (2) can be repeated for every frame in a $720x576$ $depth\ image$ sequence resulting in a 360x288 depth image sequence. The latter sequence has to be cropped to 352x288 (CIF resolution) to make it suitable for SVC encoder operation. During the up-sampling from 352x288 to 720x576, the depth video quality is slightly affected due to pixel copying at the edge to match the cropped columns, thus slightly reducing the PSNR of the up-sampled depth information. However, the overall bit rate is reduced due to down-sampling of the depth information.

The DSUS algorithm can be applied to the enhancement layer of SVC which is used to code the depth information[13]. The block diagram of the proposed method for the encoder is shown in figure 4[13][14] below at the decoder, the coded depth information from the enhancement layer is up-sampled back to its original resolution.



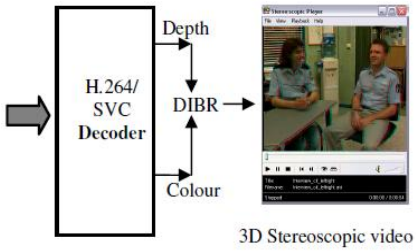


Figure 4: block diagram of H.264/SVC encoder/decoder with DSUS

IV. EVALUATION OF DEPTH CODING

The coding efficiency of common texture images is measured in the encoding bit rate and a peak signal-to-noise ratio (PSNR) value as in [14].

$$PSNR = 10\ x\ log_{10}10(\frac{255^2}{MSE}) \quad (5)$$

However, the depth image is 3D information to synthesize the virtual view, thus its quality should be evaluated in terms of rendering quality. In this letter, we measure the rendering PSNR by MSE between original image (Iorg) and rendered image (Iren) with the reconstructed depth image as in [15][16].

$$MSE_{ren} = \frac{1}{wxh}\sum_{i=0}^{w-1}(\sum_{j=0}^{h-1}||I_{org}\ (i,j) - I_{ren}(i,j)||^2) \quad (6)$$

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

Three sequences were selected for the simulations. The video frames with frame numbers 3 for Interview and Orbi depicted in figure 5 and Ballet sequence as depicted in figure 6. All the sequences include colour and depth information. The depth information is down-sampled from 720x576 spatial resolutions to CIF (352x200) resolution. The JSVM software [17] is used in the simulation for SVC. The SVC spatial scalability does not allow enhancement layer to be used to send video at lower resolution than the base layer. Therefore, a three layer configuration is used in the simulation. These sequences were acquired from Heinrich-Hertz-Institute (HHI).

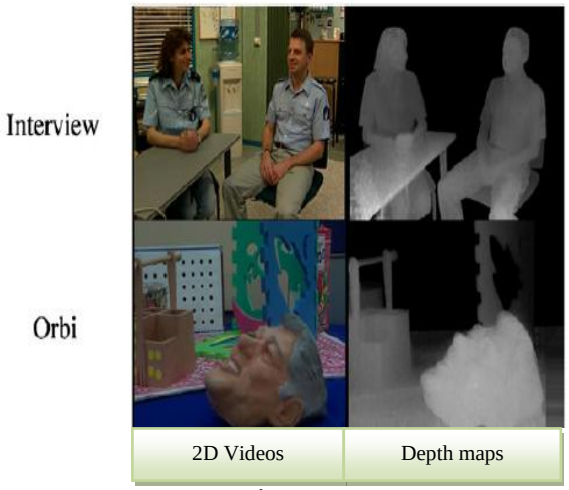


Figure 5: Depth reconstruction filter (with QP31, 33rd frame); Interview sequencerendering result without depth reconstruction filter and rendering result for Orbi sequence with depth reconstruction filter.

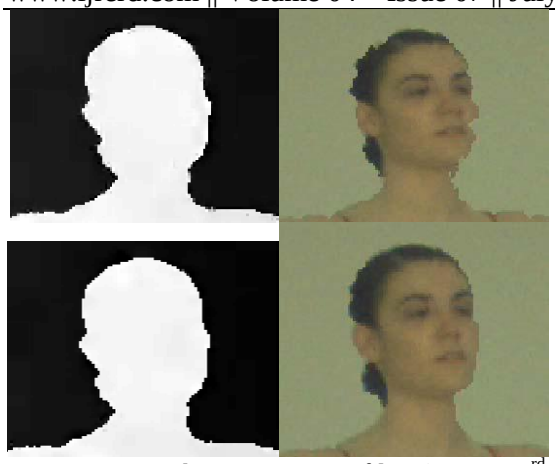


Figure 6: Depth reconstruction filter (QP31, 33rd frame) (a) without depth reconstruction filter, (b) rendering result for (a), (c) with depth reconstruction filter, (d) rendering result for (c).

For both SVC-Org and SVC-DSUS, the base layer is used to send the colour information at CIF resolution. For SVC-DSUS, the enhancement layer 1 is used to send the depth at reduced resolution. The enhancement layer 2 for both SVC-Org and SVC-DSUS is used to send the colour information at full resolution.

A. Error Free condition

The rate distortion performance of SVC-Org, SVC-DSUS are compare in an error free environment. SVC-Org is the original JSVM software[17]. The distortion is measured using the PSNR of the left and right views. An original left-and right view sequence is produced, from the original 2D texture and its associated original depth information, using the DIBR technique[18][19]. A compressed left-and-right view sequence is obtained from the 2D and depth data reconstructed at the decoder. The left and right view PSNR is obtained by comparing the original left-and-right view data with the left-and-right view data output from the decoder.

The Orbi and Interview sequences, with resolution 720x576 and 125 frame numbers are used in the error free comparison. Fixed Qp are employed to obtain the bit rates.The bit rates shown in the figures are in kbits/s. I-frames are inserted every 45 frames and only P frames are used between the I-frames. The left-and-right view PSNRs for SVC-Org and SVC-DSUS are plotted over a range of bit rates shown in figures 7 and 8. The configuration shown in table 2 is used for SVC-Org and SVC-DSUS.

It can be seen that the SVC-DSUS is less efficient that SVC-Org in terms of compression efficiency.SVC-DSUS PSNR is about 1-2 dB lower than SVC-Org for the same bit rate. This is due to the larger temporal difference between the frames used for prediction in the scheme. Larger prediction usually incurrs larger residual, and requires that larger motion vectors be coded.

Table 2: Configuration used for SVC-Org and SVC-DSUS

Encoder	**Layer**	**Spatial Resolution**	**Type**
SVC-Org	2 (enhancement)	720x576	Colour
	1 (enhancement)	720x576	Depth
	0 (base)	352x288	Colour
SVC-DSUS	2 (enhancement)	720x576	Colour
	1 (enhancement)	352x288	Depth
	0 (base)	352x288	Colour

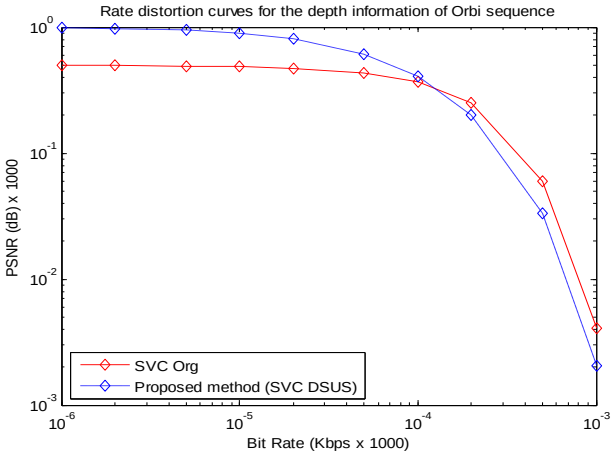


Figure 7: Rate distortion for the depth information of Orbi sequence

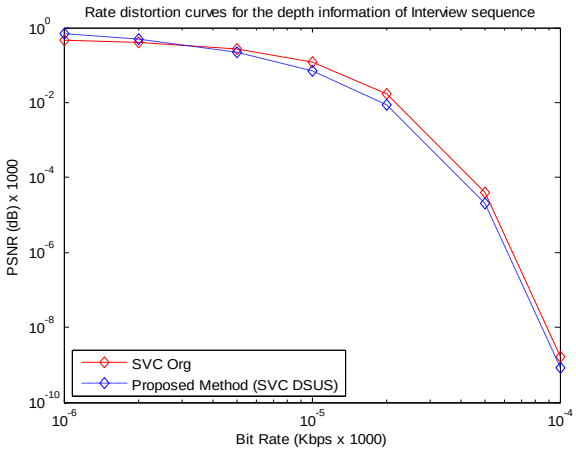


Figure 8: Rate distortion for the Interview sequence

B. Error-Proned Conditions

In this section, the performance of SVC-Org, SVC-DSUS are evaluated in an error-proned environment. The compressed 3D video is transmitted over simulated internet channels [20].

The internet channel simulator has four packet loss error pattern, namely 3%, 5%, 10% and 20%. In the simulation, the loss of one packet is assumed to mean the loss of one video frame. Frame copy error concealment is used for SVC-Org and SVC-DSUS. Both Orbi and interview sequences are used in the simulation. An I-frame is inserted every 45 frames.

Figure 9 show the rate distortion performance of the Orbi sequence for packet losses of 10%. Figure 10 show the rate distortion of the Interview sequence for packet losses of 10%. In these figures, the decoded left-and-right view PSNR for SVC-Org and SVC-DSUS is plotted. From the two figures, it can be seen that for Orbi sequence, the worst performance is given by SVC-Org at high and low bit rates. SVC-DSUS performs better with 10% packet loss rate in the low bit rate range.

For the interview sequence, SVC-DSUS performs better than SVC-Org at high error rates. SVC-Org performs worst at high error rates may be due to the up-sampling distortion in the depth image and the inability to recover quickly from errors.

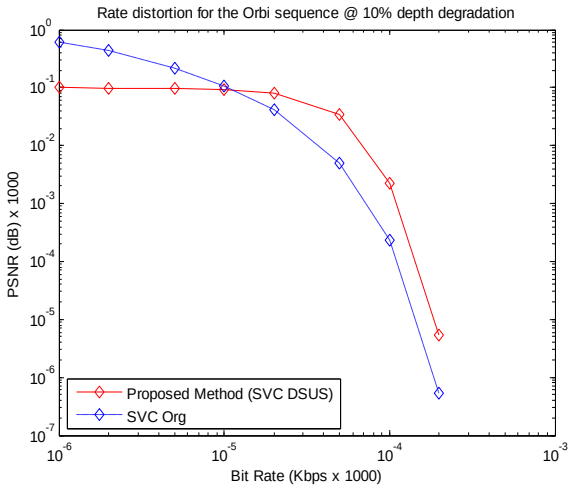


Figure 9:Rate distortion for Orbi at 10% packet loss

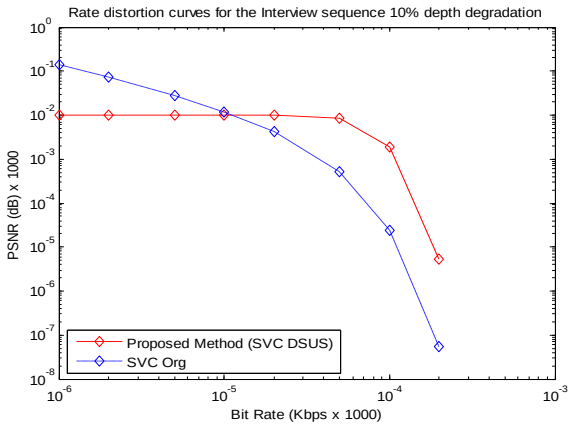


Figure 10: Rate distortion for Interview at 10% packet loss

C. Using Scalable JSVM software

Another experiments were conducted for some 3D sequences to investigate coding efficiency. Coding conditions for Ballet (Colour and Depth) were same as in above experiment for Orbi and Interview sequences. In these experiments, PSNR and bitrate are presented on R-D curves. It was decoded using B frames as in figure 1, where the view prediction with down-sampled reference pictures was applied in the proposed method. The adjacent views were encoded as I and P frames respectively. For down-sampling of reference pictures, 13 tap filters is used, which was described in the MPEG-4 VM 18 document [21]. This filter was used for generating down-sampled original picture as well.

For up-sampling as a post processing, images were generated only from decoded images of B frames. 6 tap filter{1, -5, 20, 20, -5, 1} was applied, which is the same as interpolation filter applied for generating images. Experimental results are shown in figures 11 and 12. In these figures JMVM denotes the encoding as B frame using the reference software while spatial resolution of original images is not down-sample, "down-sampled (independent)" denotes the view encoding as I view while spatial resolution of original image is down-sampled, "down-sampled (dependent)" denotes the view is encoded as B view while spatial resolution of original images and reference pictures are down-sampled, and "up-sampled (dependent)" denotes the view is up-sampled as a post processing. We see that down-sampled reference pictures, which are "down-sampled (dependent)", achieves significant bit reduction compared with independent coding and JMVM.

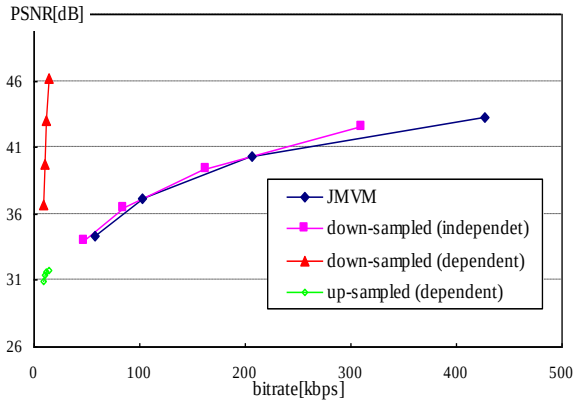


Figure 11: Coding performance of JSVM coder using several down-sampling ratios

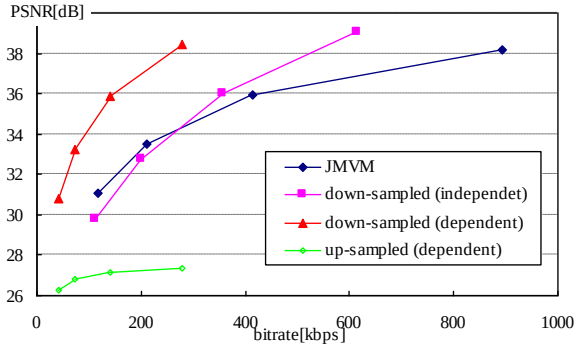


Figure 12: Coding performance of JSVM coder using several down-sampling ratios upto 1000 bit rates

VI. CONCLUSION

To improve the rate distortion, the DSUS algorithm is applied to the enhancement layer of SVC, which is used to code the depth information. The application of DSUS results in an improvement in the rate distortion performance of SVC, particularly at low bit rates. Though there is a risk of up-sampling distortion, simulation results show that the up-sampling distortion only reduces the coding efficiency under high error rates.

The paper finally investigates the performance of SVC-Org and SVC_DSUS for both Orbi and Interview sequences for stereoscopic 3D video under channel error conditions. Simulation results show that most of the time, SVC-DSUS algorithms performs better than the SVC-Org, especially in under high error-rates such as the mobile channels.

ACKNOWLEDGMENT

The work presented here was developed in CMCR lab, funded by the Petroleum Technology Development Fund (PTDF) under the Overseas Scholarship Scheme (OSS). We also thank the reviewers for their valuable contributions.

REFERENCES

[1] A. Sadka, "Compressed Video Communications" John Willey & Sons Ltd, 2002.
[2] O. Schreer, P. Kauff, and T. Sikora, *3D video communication*, West Sussex, John Wiley & Son Ltd., pp. 29-37, 2005.
[3] Ekmekcioglu, E., Worrall, S.T. & Kondoz, A.M., 2008. Utilisation of downsampling for arbitrary views in multi-view video coding. *Electronics Letters*, 44(5), 8-9.
[4] Oh, K. et al., 2009. Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3D Video. *IEEE Signal Processing Letters*.
[5] Shi, Z. et al., Adaptive DCT-Domain Down-Sampling and Learning Based Mapping for Low Bit-Rate Image Compression. *Image (Rochester, N.Y.)*, 222-231.
[6] http://ftp3.itu.org/av-arch/jvt-site/2007_04_SanJose/
[7] Schwarz, H., Marpe, D. & Wiegand, T., Scalable Extension of H.264 / AVC. *Current*.
[8] Wiegand, T. & Sullivan, G.J., 2007. The H.264/AVC Video Coding Standard. *IEEE Signal Processing Magazine*, (August 1999), 148-153.

[9] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression", IEEE Trans. Image Processing, vol.12, no 12, pp. 1132-1144, September 2003.
[10] Ekmekcioglu, E., Worrall, S.T. & Kondoz, A.M., 2008. Bit-rate adaptive downsampling for the coding of multi-view video with depth information. *Network*, 2008-2011.
[11] Ekmekcioglu, E, Vladan V, S.W., Efficient Edge, Motion and Depth-Range Adaptive Processing for Enhancement of Multiview Depth Maps sequences. *Processing*, 2-5.
[12] Yasakethu, S.L., Fernando, W.A. & Kondoz, A.M., 2009. Rate Controlling in Off Line 3D Video Coding Using Evolution Strategy. *Strategy*, 55(1), 150-157.
[13] Karim, H.A., Worrall, S. & Kondoz, A.M., Reduced Resolution Depth Compression for Scalable 3D Video Coding. *Engineering and Technology*, 765-770.
[14] H. Karim, a. Sali, S. Worrall, A. Sadka, and a. Kondoz, "Multiple description video coding for stereoscopic 3D," *IEEE Transactions on Consumer Electronics*, vol. 55, 2009, pp. 2048-2056.
[15] W. Chen, Y. Chang, S. Lin, L. Ding, and L. Chen, "Efficient depth image based rendering with edge dependent depth filter and," *interpolation, in: Proceedings of International Conference on Multimedia and Expo*, 2005, pp. 1314-1317.
[16] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, a. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, 2007, pp. 217-234.
[17] http://freedownloadbooks.net/jsvm-software-download-pdf.html
[18] Fehn, C., A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR). *Image (Rochester, N.Y.)*.
[19] A. Redert, M. de Beeck, C. Fehn, W. Ijsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, and P. Surman, "Advanced three-dimensional television system technologies," *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*, 2002, pp. 313-319.
[20] http://www.shunra.com/internet_simulation_testing
[21] http://focus.ti.com/docs/toolsw/folders/print/tmdmpeg4e.html

Umar Abubakar Sadiq received B. Eng. Degree in Electrical/Electronics from Ahmadu Bello University, Zaria in 1995. He got a scholarship to read Master of Science in Mobile and Satellite Communications from University of Surrey, UK and graduated with 2/1 in 2004. Mr. Umar is with the National Space Research & Development Agency, Abuja, Nigeria and completed his PhD programme in Centre for Media Communication Research (CMCR), department of Electronic & Computer Engineering in Brunel University, West London. His current research interest is in the area of 3D video communications over wireless networks, video and image processing, video compression, Moblie and Satellite Communication, Spectrum Management etc.

Abdul H. Sadka (Senior Member IEEE) received the B.Eng. degree in computer and communications engineering, the M.Sc. degree in computer engineering, and the Ph.D. degree in electrical and engineering, in 1990, 1993, and 1997, respectively. Professor A. H. Sadka is the head of Electronic and Computer Engineering and the director of the centre for Media Communications Research in Brunel University, UK, with almost 15-year experience in academic leadership and excellence. He is an internationally renowned expert in visual media processing and communications with an extensive track record of scientific achievements and peer recognised research excellence. He has managed so far to attract over £2M worth of research grants and contracts in his capacity as principal investigator. He has published widely in international journals and conferences and is the author of a highly regarded book on "Compressed Video Communications" published by Wiley in 2002. He holds 3 patents in the video transport and compression area. He acts as scientific advisor and consultant to several key companies in the international Telecommunications sector and is the founder and managing director of VIDCOM Ltd.