ENHANCEMENT METHOD FOR MULTIPLE DESCRIPTION DECODING OF DEPTH MAPS SUBJECT TO RANDOM LOSS

P. Correia[1,2], *S. Marcelino*[1,4], *P. Assuncao*[1,3], *S. Faria*[1,3] *S. Soares*[4,5], *C. Pagliari*[6] *and E. da Silva*[7]

[1] Instituto de Telecomunicações; [2]Polytechnic Institute of Tomar,
[3]Polytechnic Institute of Leiria; [4] UTAD, Engineering Department;[5] IEETA UA CAmpus, Portugal;
[6]DEE,Instituto Militar de Engenharia, [7]PEE/COPPE/DEL/Poli,Univ. Federal do Rio de Janeiro,Brazil

e-mails:{pcorreia,stam}@co.it.pt, {pedro.assuncao, sergio.faria}@ipleiria.pt, salblues@utad.pt, carla@ime.eb.br, eduardo@smt.ufrj.br

ABSTRACT

This paper proposes a method to improve the quality of depth maps transmitted in multiple descriptions through multipath error prone networks. Whenever a single description is lost, the remaining ones are still able to provide a coarsely decoded version of the depth map. While in multiple description video, such coarse decoding is still acceptable for display, in the case of depth maps the additional decoding distortion propagates through the corresponding view synthesis. The proposed method is capable of enhancing low quality depth maps decoded from one single description, based on geometric information available in coarsely decoded slices, combined with higher quality depth values in adjacent slices decoded from both descriptions. In comparison with existing MDC decoders, the enhancement method achieves quality gains in synthesised views up to 1.69dB for packet loss rates of 10%.

Index Terms **—** Multiple description decoding, depth map enhancement, depth map packet loss.

1. INTRODUCTION

Services and applications using multiview images and video have been receiving increasing attention in the last decade and this is expected to continue as a fast developing technology field in the forthcoming years. As a consequence, compressed 3D and multiview visual content will be a major contributor to the overall multimedia traffic delivered through various types of channels. Nowadays, there are several types of coded representations of 3D and multiview image and video content, where the most common are the multiview video coding (MVC) and multiview video-plus-depth (MVD) [1] [2]. In MVC, stereoscopic viewing is directly obtained from two adjacent views, while in MVD each colour image is associated with a corresponding depth map, which can be used to synthesise other views.

It is known that the impact of transmission errors or data loss is greater in 3D image and video communications than in traditional 2D communications, because the perceived quality of experience (QoE) is highly sensitive to a wider variety of quality factors [3]. In this context, multiple description coding (MDC) is an efficient approach to deal with error-prone multipath networks [4]. In MDC the source signal is encoded in two or more independently decodable streams (i.e., descriptions) which comprise correlated representations of the same source [5]. The great advantage of MDC is that a minimum signal quality is almost guaranteed because the probability of simultaneous loss of all descriptions is low in comparison with single stream coding and transmission, i.e., single description coding (SDC). However, the use of MDC reduces the coding efficiency. Yet, the improved robustness compensate for the loss of coding efficiency and the overall quality obtained after decoding MDC streams is better than SDC, especially for higher packet loss rates [6]. This paper extends the MDC approach to depth maps, in the context of multiview video communications over multiple lossy channels.

In the past, some MDC methods have been proposed in 3D and multiview video communications. In [7], MDC is proposed for stereoscopic video based on spatial and temporal scaling. Temporal sub-sampling is also used in [8] based on a 3D even and odd MDC scheme with adaptive redundancy added to frames with motion activity higher than a threshold. An MDC scheme based on scalable coding was proposed in [9], where a redundant version of the enhanced layer is encoded as different description. In [10], an MDC scheme is presented using video plus depth coding, with viewpoint synthesis. Scalable coding is used and each of the two viewpoints is encoded with two spatial layers and two temporal layers. In addition to the scalable layers, a redundant stream for the base layer is generated with redundant encoding of only the most significant foreground objects. A simulcast encoding scheme of temporal sub-sampled versions of each view is proposed in [11], using depth-image-based rendering (DIBR) to synthesise the missing view and motion compensation concealment for occluded areas. In [12] a multiview MDC scheme based on spatial down sampling is proposed using multiview video coding amendment (MVC) of the H.264/AVC. All the previous works emphasize the effectiveness of MDC for 3D image and video signals in multipath video communications and its higher performance in comparison with traditional SDC. However, delivery of depth maps encoded as multiple descriptions and transmitted though diverse error prone channels was not addressed.

This paper deals with the problem of using MDC for multipath delivery of depth maps. Whenever any description is not available for decoding (e.g., due to packet loss), the resulting coarsely reconstructed values produce a significant amount of distortion into the synthesised images. Given the particular nature of depth information, i.e., smooth gray-level areas with reasonably well defined boundaries, a method to enhance the accuracy of depth maps decoded from one single description is proposed to improve the quality of synthesised images. Geometric information extracted from object boundaries of the depth map is used to compute new depth values, according to the relative spatial position of neighbour values.

This paper is organized as follows. Section 2 explains the basics of MDC applied to depth maps and Section 3 describes the enhancement method proposed for MDC decoding. Experimental results are presented and discussed in Section 4 and Section 5 concludes the paper.

978-1-4799-4758-4/14/$31.00 ©2014 IEEE

2. MULTIPLE DESCRIPTION CODING OF DEPTH MAPS

The MDC method used in this work to encode depth maps is based on Multiple Description Scalar Quantisation (MDSQ) to generate two independent descriptions from the same source signal, based on two functions: *scalar quantisation* and *index assignment* [13]. Scalar quantisation is generally applied to transform coefficients x, such that an index i_0 is produced for each one. Then the index assignment process is based on an *index assignment matrix* (Table 1) to produce a side index pair (i_1, i_2) from each central index i_0. Each side index corresponds to a quantisation cell containing several possible central indices. For instance, in Table 1, side index $i_1 = 2$ corresponds to a cell with central indices $i_0 \in \{4, 6, 7\}$. Considering two descriptions with rate-distortion pairs $(R1, D1)$ and $(R2, D2)$, these two descriptions are generated from a central quantiser with rate-distortion pair $(R0, D0)$, using index assignment. This is equivalent to generating the two descriptions using coarser side quantisers. In general, MDC algorithms use balanced descriptions where the respective rates and distortions are approximately the same, i.e., $R1 \approx R2$ and $D1 \approx D2$.

For a given source signal, an index assignment is defined to give the best row-column assignment of central quantisation indices along the matrix diagonal. For a selected set of side index pairs this should result in small spread in each cell. For a given index assignment matrix, the number of cells (N) is determined by the number of different side quantisation indices defined for that matrix. The central and side distortions $D0$ and $D1$ or $D2$, are defined as the distortion obtained from decoding both descriptions or only one of them, respectively. The side distortion is obviously higher than the central one and corresponds to the case where any single description is lost in the transmission path.

Both the side distortion and the amount of redundancy depend on the index spread (k), which represents the number of diagonals above (or below) the main diagonal in the index assignment matrix. Note that central distortion $D0$ does not change with k for a given fixed N whereas side distortions $D1$, $D2$ increase with k. In the example of Table 1, $k = 1$ which results in 3 diagonals.

Table 1: Index assignment matrix ,k=1

		i2 (Description 2)								
		-4	-3	-2	-1	0	1	2	3	4
i1 (Descr. 1)	-2		-8	-6	-7					
	-1			-4	-3	-1				
	0				-2	0	1			
	1					2	3	5		
	2						4	6	7	

The MDC encoder includes the MDSQ module to produce two descriptions and the corresponding streams for transmission over different paths. The MDC encoding architecture uses both intra predicted and motion compensated predicted slices [6]. The headers, prediction modes and motion vectors are duplicated in both descriptions. At the decoder, if both descriptions are available, then an inverse index assignment process is used to restore a unique central index i_0 to be inverse quantised and transformed. In the case where one description is not available for decoding due to transmission errors/losses, the other one is decoded using an equivalent coarser quantisation stepsize, which leads to a lower quality decoded slice.

Although other MDC methods exist with equivalent coding efficiency, the proposed one has the advantage of not changing the resolution (either spatial or temporal) of the depth map. The proposed enhancement method takes advantage from the higher quality of the information received in the neighboring areas of the lost slices.

3. ENHANCEMENT METHOD FOR MDC DEPTH MAPS

As can be seen in Figure 1, the proposed method comprises four processing steps to enhance the quality of MDC depth maps decoded from only one MDSQ description. Description losses are supposed to occur randomly at slice level, which means that not all slices of a depth map are necessarily lost at the same time. If both descriptions are lost, then a loss concealment method is used at the decoder, such as spatial interpolation, motion-copy or region-copy. In this case, the proposed enhancement process is bypassed.

The first step of the enhancement process is to extract the contour information from depth slices affected by the loss of one description. Despite the low quality of slices decoded from single descriptions, in general, for the main objects in the 3D scene, it is still possible to extract depth contours with enough accuracy to be used by the enhancement process. The Canny Edge extraction algorithm is used to extract the depth map contours [14].

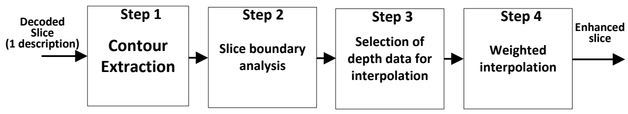


Figure 1: Enhancement process for MDC depth maps

The second step consists in analysing the slice boundaries in order to find out whether a slice was decoded from one single description (low quality) and any of its neighbours from two descriptions (high quality). If there is a consistent difference between spatially adjacent slices, then an edge, coincident with their common borders, should be found. In this case, the coarsely decoded depth values are enhanced through steps 3 and 4, as described in the Subsections 3.1 and 3.2.

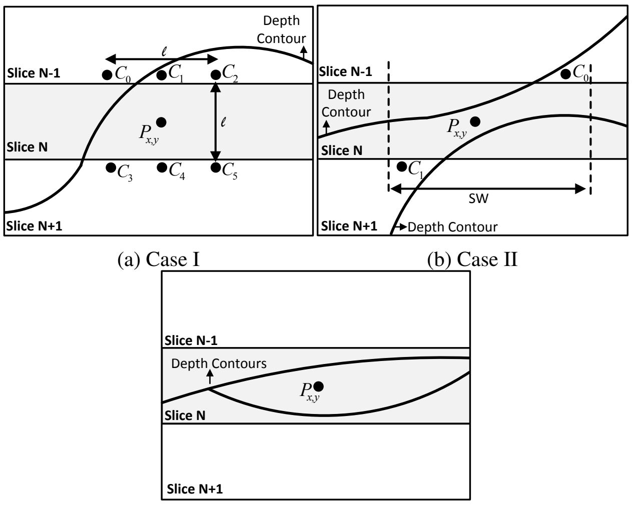
3.1. Depth data selection

The object contours found in a single description slice are used to select the appropriate data to enhance each coarsely decoded depth value. Three different cases can be used, according to the available relevant data (Figure 2).

Case I In the first case, shown in Figure 2a, if accurate neighbour depth values (i.e., a slice decoded from both descriptions) are available, then these are used to interpolate the coarse depth value ($P_{(x,y)}$). Up to six candidate depth values can be used: C_0 to C_2, from top slice N-1, and C_3 to C_5 from bottom slice N+1. If a certain candidate value is beyond the contour, then it is not used for interpolation because it belongs to a quite different depth level. If all six candidates are beyond the contour, then the coarse value fits into one of the next two cases.

Case II In the second case, shown in Figure 2b, interpolation of the coarse value uses up to two neighbour candidates, one from the top slice (C_0) and another from the bottom slice (C_1). These values are found by searching inside the region delimited by a predefined search window (SW). If this search is not successful in finding neighbour depth values belonging to the same region, i.e. delimited by the extracted countours, then coarse depth value belongs to the third case.

Case III In the third case shown in Figure 2c, the coarse depth values usually belong to small object contours inside the coarsely decoded slice. Therefore, using neighbour data to enhance these depth values does not bring any guaranteed benefit. In this case the coarse depth value is not modified.



(a) Case I (b) Case II

(c) Case III

Figure 2: Use of depth contours

3.2. Weighted interpolation

After selecting the neighbour depth values (C_i), as described above, weighted interpolation is used to improve the accuracy of the coarse depth values. The following expression is used:

$$P^{'}_{(x,y)} = \frac{P_{(x,y)} * \frac{d_P}{0.5*l} + \sum_{i=1}^{N} C_i \times [1 - \frac{d_i}{MAX_{distance}}]}{\frac{d_P}{0.5*l} + \sum_{i=1}^{N} (\frac{d_i}{MAX_{distance}})}, \quad (1)$$

where $P_{(x,y)}$ and $P^{'}(x,y)$ are the coarse and enhanced depth values, respectively, with coordinates x and y. l is the slice height, d_p is the distance between $P(x,y)$ and the top or bottom limits of the slice being enhanced (the smallest of these two distances is chosen for d_p), N is the number of values used for interpolation, C_i is depth value i, d_i is the distance between $P_{(x,y)}$ and C_i. Finally, $MAX_{distance}$ is the maximum distance between depth values used in the interpolation.

4. EXPERIMENTAL RESULTS

The performance of the proposed method was evaluated by using two views of three MVD sequences with same resolution (1024×768) were used: Ballet (view0,view2), Book Arrival (view8, view10) and Break Dancers (view0, view2). Each sequence has different characteristics: Book Arrival has moderate object motion and complex depth structure, Break dancers exhibits many objects in the scene (at different depths), and Ballet is less complex with large objects. The reference software View Synthesis Reference Software (VSRS 3.5) was used to synthesise the intermediate view. The contours of depth maps were determined using the Canny algorithm implementation in *OpenCV 2.1* [15]. The low and high thresholds for the *Canny* algorithm were computed based on the $mean$ of the image under analysis, i.e., $th_{low} = 0.5 * mean$ and $th_{high} = 1.5 * mean$, respectively.

The colour images and the corresponding depth maps used in the simulations were encoded at fixed QP=28 using H.264/AVC, using the reference software. One hundred frames were encoded using an IDR period of 12, and a GOP structure *IBPBP* with two reference frames. Slice mode was also used and each depth map was divided into slices with height of 64 pixel and width equal to the horizontal resolution. Each slice is packetised into one single packet. Two MDC balanced descriptions were used in the simulations, generated by an MDC encoder with 3 diagonals in index assignment matrices. The redundancy of the MDC of the depth maps scheme is inline with the obtained in the traditional MDC for texture video [6].

In these experiments two independent network paths are simulated to deliver each description, each one subject to equal packet loss rate (PLR). A random packet loss generator with uniform distribution was used to drop packets according to the required PLR. Transmission of each sequence of depth maps was simulated 10 times under the same PLR. Colour views are assumed to be transmitted without losses. As mentioned before, when both descriptions of any slice are simultaneously lost, a classic error concealment algorithm is used, either i) Spatial weighted interpolation using the high quality depth values from error free neighbouring regions; ii) Motion-copy or iii) Region Copy, as defined in the reference software JM17. In the case where only one description is lost, the proposed method is used to enhance the quality of the depth map decoded from one single description with the inherent higher distortion.

The algorithm was evaluated by computing the PSNR of synthesised views obtained from MDC depth maps transmitted through different channels with random losses of 2%, 5%, and 10%. The objective quality is evaluated using the virtual view synthesised with the uncorrupted depth map instead of the actual view, to avoid the influence of the DIBR algorithm in the results. The PSNR of the synthesised views obtained from enhanced and not-enhanced depth maps is shown in Table 2, for the three concealment methods. Note that concealment is only used when both descriptions are simultaneously lost. The quality difference is shown in column ΔPSNR. For reference, the PSNR of synthesised views obtained from error-free depth maps is also shown, i.e., PLR = 0%.

The simulation results show that the proposed method consistently enhances the reconstructed depth maps with positive impact on the quality of synthesised views. The higher gains occur in sequences with smooth areas and large objects, at different depth levels. Since lost descriptions correspond to an entire slice, which is a large area in the depth map, without the enhancement process, very often the synthesised images present poorer quality. Moreover, the distortion effect of inaccurate depth slices propagates throughout the GOP, contributing for additional degradation on the corresponding synthesised images. The maximum improvement in Table 2 is 1.69dB, obtained in *Ballet* sequence for the worst case of 10% error loss. The proposed method is also able to achieve consistent quality improvement for different types of sequences and concealment methods.

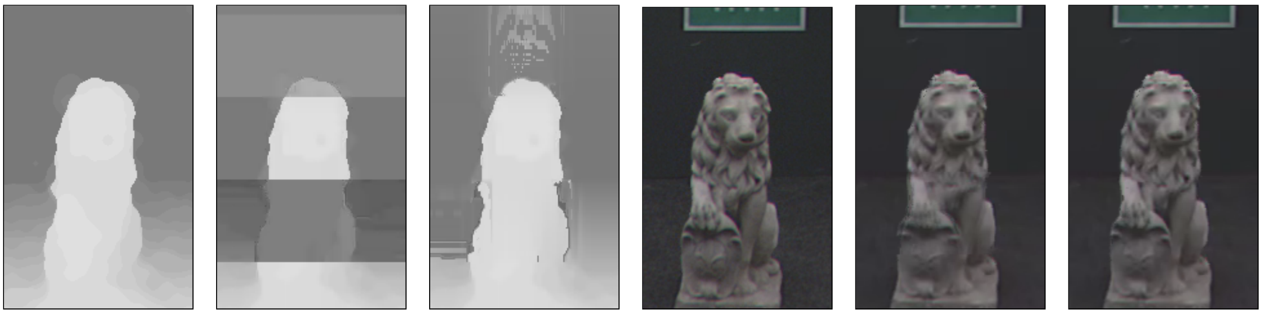
These objective results can be subjectively confirmed, for instance in Figure 3, where the synthesised images are shown along with the respective depth maps. The example of Figure 3 shows a region detail of the original depth map (Figure 3a), the coarsely decoded depth map with 10% of loss (Figure 3b), and the enhanced depth map using the proposed method (Figure 3c). The corresponding synthesised views using the original depth map is shown in Figure 3d, the synthesised view using the non-enhanced depth map is shown in Figure 3e and the synthesised view using the proposed enhancement method is shown in Figure 3f. These detailed regions show that when there are large depth variations, the proposed method is able to reconstruct the lost regions with very high accuracy. In the presence of thin objects fully included in the coarse area, it is much more difficult to achieve improve the depth map quality.

5. CONCLUSIONS

This paper presents a method to enhance MDC depth maps decoded from one single description. The results show that the pro-

Table 2: Average PSNR-Y of synthesised views

		Not Enhanced (a)			Enhanced (b)			ΔPSNR: (b)-(a)		
PLR	0%	2%	5%	10%	2%	5%	10%	2%	5%	10%
Concealment method:	**Spatial interpolation**									
Ballet	35.99	34.69	33.61	32.25	35.24	34.77	33.90	0.55	1.16	1.65
Book Arrival	41.02	40.26	39.33	37.81	40.58	40.09	38.91	0.32	0.76	1.1
Break Dancers	36.34	34.94	34.06	32.53	35.18	34.39	33.32	0.24	0.33	0.79
Concealment method:	**Motion copy**									
Ballet	35.99	34.69	33.61	32.26	35.24	34.77	33.94	0.55	1.16	1.68
Book Arrival	41.02	40.26	39.33	38.87	40.58	40.11	38.99	0.32	0.78	1.1
Break Dancers	36.34	34.94	34.05	32.48	35.18	34.39	33.29	0.24	0.34	0.81
Concealment method:	**Region Copy**									
Ballet	35.99	34.69	33.61	32.25	35.24	34.77	33.94	0.55	1.16	1.69
Book Arrival	41.02	40.26	39.33	37.87	40.58	40.10	38.99	0.32	0.77	1.12
Break Dancers	36.34	34.94	34.05	32.49	35.18	34.39	33.90	0.24	0.34	0.81



(a) Original (b) No enhancement (c) Enhancement (d) Original (e) No enhancement (f) Enhancement

Figure 3: MDC depth maps and respective synthesised views at 10% of packet loss (frame 84).

posed method is able to efficiently enhance the lower quality depth maps by using the contours as region boundaries for weighted interpolation in the coarsely decoded regions. The synthesised images, using such enhanced depth maps, consistently achieve higher quality than those not enhanced. Overall, the superior performance of the proposed method demonstrates that the accuracy of reconstructed MDC depth maps can be increased to improve the quality of synthesised views.

6. ACKNOWLEDGMENT

This work was supported by EU COST Action IC1105 (3D - ConTourNet) and FCT, Portugal: Grant SFRH/BD/64988/2009, Project PTDC/EEA-TEL 114487/2009.

7. REFERENCES

[1] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, 2009.

[2] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, April 2011.

[3] C. T E R Hewage, S.T. Worrall, S. Dogan, S. Villette, and A.M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 2, pp. 304–318, April 2009.

[4] P. Frossard, J.C. De Martin, and M.R. Civanlar, "Media streaming with network diversity," *Proceedings of the IEEE*, vol. 96, no. 1, pp. 39–53, Jan 2008.

[5] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[6] P. Correia, P. Assuncao, and V. Silva, "Multiple description of coded video for path diversity streaming adaptation," *IEEE Transactions on Multimedia*, vol. 14, no. 2-3, pp. 923–935, June 2012.

[7] A. Norkin, A. Gotchev, K. Egiazarian, and J. Astola, "Two stage multiple description image coders: Analysis and comparative study," *Signal Processing: Image Communications*, vol. 11, pp. 609–625, 2006.

[8] H. A. Karim, A. Sali, S. Worrall, Abduk H. Sadka, and A.M. Kondoz, "Multiple description video coding for stereoscopic 3D," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 2048–2056, November 2009.

[9] M.B.Dissanayake, D.V.S.X. De Silva, S.T.Worral, and W.A.C. Fernando, "Error resilience technique for multi-view coding using redundant disparity vectors," in *IEEE International Conference On Multimedia and Expo, ICME, 2010*, 2010, pp. 1712–1717.

[10] E. Ekmekcioglu, B. Günel, M. Dissanayake, S. T. Worral, and A. M. Kondoz, "A scalable multi-view audiovisual entertainment framework with content-aware distribuition," in *IEEE International Conference On Image Processing, ICIP, 2010*, Sept. 2010, pp. 2401–2404.

[11] Z. Liu, G. Cheung, J. Chakareski, and Yusheng Ji, "Multiple description coding of free viewpoint video for multi-path network streaming," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 2150–2155.

[12] X. Wang and C. Cai, "Mode duplication based multiview multiple description video coding," in *Data Compression Conference (DCC)*, March 2013, pp. 527–527.

[13] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions On Information Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[14] J. Canny, "A Computational Approach to Edge Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679 –698, November 1986.

[15] OpenCV-2.1.0, *http://opencv.itseez.com/*.