

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2023-02-07

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Conti, C., Nunes, P. & Soares, L. D. (2014). Impact of packet losses in scalable 3D holoscopic video coding. In Schelkens, P., Ebrahimi, T., Cristóbal, G., Truchetet, F., and Saarikko, P. (Ed.), SPIE Photonics Europe - Optics, Photonics, and Digital Technologies for Multimedia Applications III. Bruxelles: Society of Photo-Optical Instrumentation Engineers.

Further information on publisher's website:

10.1117/12.2052790

Publisher's copyright statement:

This is the peer reviewed version of the following article: Conti, C., Nunes, P. & Soares, L. D. (2014). Impact of packet losses in scalable 3D holoscopic video coding. In Schelkens, P., Ebrahimi, T., Cristóbal, G., Truchetet, F., and Saarikko, P. (Ed.), SPIE Photonics Europe - Optics, Photonics, and Digital Technologies for Multimedia Applications III. Bruxelles: Society of Photo-Optical Instrumentation Engineers., which has been published in final form at <https://dx.doi.org/10.1117/12.2052790>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Impact of packet losses in scalable 3D holoscopic video coding

Caroline Conti^{*a,b}, Paulo Nunes^{a,b}, Luís Ducla Soares^{a,b}

^aISCTE – University Institute of Lisbon, Av. das Forças Armadas, 1649-026 Lisbon, Portugal;

^bInstituto de Telecomunicações, Av. Rovisco Pais 1, 1049-001 Lisbon.

ABSTRACT

Holoscopic imaging became a prospective glassless 3D technology to provide more natural 3D viewing experiences to the end user. Additionally, holoscopic systems also allow new post-production degrees of freedom, such as controlling the plane of focus or the viewing angle presented to the user. However, to successfully introduce this technology into the consumer market, a display scalable coding approach is essential to achieve backward compatibility with legacy 2D and 3D displays. Moreover, to effectively transmit 3D holoscopic content over error-prone networks, e.g., wireless networks or the Internet, error resilience techniques are required to mitigate the impact of data impairments in the user quality perception. Therefore, it is essential to deeply understand the impact of packet losses in terms of decoding video quality for the specific case of 3D holoscopic content, notably when a scalable approach is used.

In this context, this paper studies the impact of packet losses when using a three-layer display scalable 3D holoscopic video coding architecture previously proposed, where each layer represents a different level of display scalability (i.e., L0 - 2D, L1 - stereo or multiview, and L2 - full 3D holoscopic). For this, a simple error concealment algorithm is used, which makes use of inter-layer redundancy between multiview and 3D holoscopic content and the inherent correlation of the 3D holoscopic content to estimate lost data. Furthermore, a study of the influence of 2D views generation parameters used in lower layers on the performance of the used error concealment algorithm is also presented.

Keywords: Display scalability, HEVC, integral imaging, 3D holoscopic video coding, Error Resilience Techniques

1. INTRODUCTION

Three dimensional (3D) holoscopic imaging, also known as integral imaging and plenoptic imaging is a light field imaging technology that has been attracting the attention of research community and camera manufacturing industry for providing richer two-dimensional (2D) image capturing systems [1], single-camera 3D imaging and more immersive 3D viewing systems [2]. Moreover, the recent advances in lenses manufacturing and sensor resolutions allowed the development of practical imaging devices for this type of content [3][4], promising to become a popular technology in near future.

This is possible due to the used optical structure, where a micro-lens array is overlaid on the camera sensor for content acquisition and on a conventional flat panel display for content display. At the capturing side, the 3D holoscopic imaging system is able to capture all the spatial and angular information about the scene, since each micro-lens captures a low resolution portion of scene at slightly different angles to its neighbor. At the display side, the captured light field representing the original scene can be reconstructed throughout the viewing zone with continuous motion parallax in all directions (horizontal and vertical). This allows the user to have a truly 3D viewing experience and significantly reduces eyestrain compared to current 3D stereoscopic and multiview displays.

However, in order to progressively introduce this technology into the consumer market and to efficiently deliver 3D holoscopic content to end-users, a crucial requirement is backward compatibility with legacy 2D and 3D devices. Hence, to enable 3D holoscopic content to be delivered and presented on legacy displays, a 3D holoscopic scalable coding approach is required, where by decoding only the adequate subsets of the scalable bitstream, 2D or 3D compatible video decoders can present an appropriate version of the 3D holoscopic content.

Moreover, following the current forecasts indicating that over two-thirds of the world's mobile data traffic will be video by 2018 [5], it should be envisaged to efficiently provide 3D holoscopic video services in such type of error prone environments. To guarantee this, error resilience techniques in the encoding and decoding side are needed to mitigate the impact of data impairments in the user quality perception. The design of an appropriate error resilience technique typically takes into account the type of network (i.e., error characteristics of the network being used for transmission) and also the type of content (i.e., inherent characteristics of the content) being transmitted. In this sense, due to the

*caroline.conti@it.lx.pt; phone +351 218418164; www.it.pt

different nature of acquisition system and, consequently, the different type of correlation in the 3D holoscopic content, when compared to the conventional 2D and 3D multiview contents, the set of factors which could affect the performance of an error control algorithm may also differ. Hence, it is essential to deeply understand the impact of packet losses in terms of decoding video quality for the specific case of 3D holoscopic content, notably when a scalable approach is used.

To the best of the authors' knowledge, the proposal of error resilience techniques suitable for 3D holoscopic content has not been addressed in the literature yet. Therefore, this paper proposes to start the discussion in this issue and presents a study of the influence of packet losses in scalable 3D holoscopic content coding. For this, a three layer scalable 3D holoscopic video coding architecture – previously proposed by the authors in [6] – is considered, where each layer represents a different level of display scalability (i.e., L0 - 2D, L1 - stereo or multiview, and L2 - full 3D holoscopic). Based on this coding architecture, a simple error concealment algorithm is proposed, which derives from the previously proposed inter-layer prediction method (in [6]) to estimate the lost data. Finally, an analysis of the influence of some meaningful parameters in the proposed coding architecture (e.g., 2D view generation parameters used in lower layers) on the performance of the used error concealment algorithm is also presented.

The remainder of this paper is organized as follows: Section 2 presents the used scalable architecture for 3D holoscopic video coding, as well as the process to generate the content for each coding layer; Section 3 reviews the inter-layer prediction scheme, in order to better understand the proposed error concealment algorithm; Section 4 some meaningful factors which affect the inter-layer prediction accuracy, and presents the proposed error concealment algorithm; Section 5 presents the considered test conditions and studies the influence of packet losses on the accuracy of inter-layer prediction; and finally, Section 6 concludes the paper and presents some future work.

2. 3D HOLOSCOPIC SCALABLE CODING ARCHITECTURE

A multi-layer architecture for 3D holoscopic coding was proposed in [6], as illustrated in Figure 1, so as to support display scalability. In this architecture, each hierarchical layer represents different levels of display scalability:

- *Base Layer*: The base layer represents a single 2D view, which can be used to deliver a 2D version of the 3D holoscopic content to 2D displays;
- *First Enhancement Layer*: This layer represents the necessary information to obtain an additional view (representing a stereo pair) or various additional views (representing multiview content). It intends to allow stereoscopic displays or multiview displays to support 3D holoscopic content;
- *Second Enhancement Layer*: This layer represents the additional data needed to support full 3D holoscopic video content.

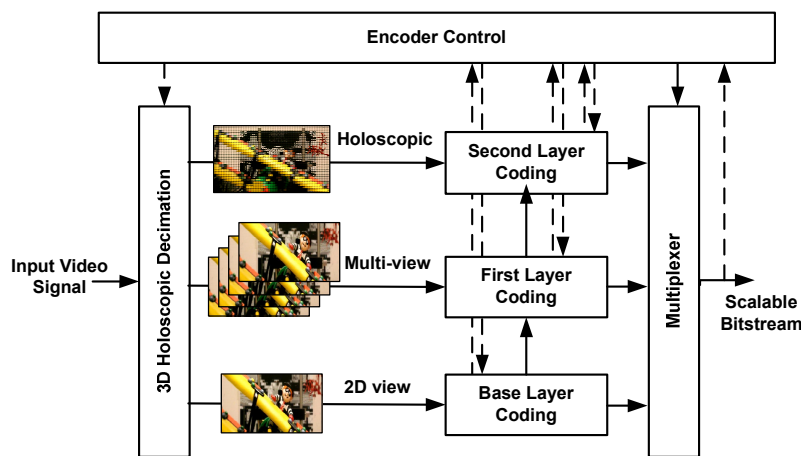


Figure 1 Scalable 3D holoscopic video coding architecture using three hierarchical layers.

It is important to notice that, if a MVC approach is used for the *Base Layer* and *First Enhancement Layer*, display scalability is intrinsically supported and a portion of the bitstream can be accessed in order to output a subset of encoded views. Thus, it is possible to generalize the aforementioned architecture by distinguishing more layers, for instance, one for each different additional view.

Therefore, the coding information flow in the proposed scalable 3D holoscopic coding architecture (for one access unit) is the following:

- i) In the *Base Layer*, the 2D views are coded with a suitable standard 2D video encoder and the reconstructed frames are kept for coding the upper layers;
- ii) Between the *Base Layer* and the *First Enhancement Layer*, an inter-layer prediction mechanism exploits the existing inter-view correlation to improve the coding efficiency, for instance, any of the schemes in [7]. Similarly, within the *First Enhancement Layer* an inter-view prediction is also used and the encoded and reconstructed data is fed to the *Second Enhancement Layer*;
- iii) Between the *First* and the *Second Enhancement Layers*, inter-layer prediction (see Section 3) exploits the optical-geometric relation between the multiview content and the 3D holoscopic content to take advantage of as much redundant information as possible. Moreover, within the *Second Enhancement Layer*, as proposed previously by the authors [8], a self-similarity compensated prediction is used to exploits the inherent redundancy of the holoscopic content.

Furthermore, to be able to display 3D holoscopic content on 2D and 3D multiview displays, it is necessary to produce adequate versions of the content for the *Base Layer* and the *First Enhancement Layer*. This means to generate various 2D views from 3D holoscopic content, as represented by the 3D holoscopic decimation block in Figure 1.

In this context, one of the advantages of employing a 3D holoscopic imaging capturing system is that it opens new degrees of freedom in terms post-production tools. Notably, a single 3D holoscopic image is able to provide 2D view images where focus, exposure, chosen perspective and even depth of field [9] are adjustable a posteriori.

Several algorithms to generate 2D view images from a 3D holoscopic image have been proposed in the literature, mainly in the context of richer 2D image capturing systems [10][11]. In this paper, two of these algorithms, referred to as *Basic Rendering* and *Weighted Blending* [1] are considered to generate the content for the first two hierarchical layer of the scalable 3D holoscopic coding scheme.

Briefly, the both *Basic Rendering* and *Weighted Blending* algorithms are based on the fact that since each micro-image can be seen as a low resolution view of the scene, it is possible to choose suitable portions (patches) from each micro-image to properly compose a 2D view image. Then, as explained in [1], the process of generating a 2D view image can be controlled by the following two main parameters:

- i) *Size of the patch*: Since a micro-image is captured by the corresponding micro-lens in perspective projection geometry, placing two objects of the same (real) size closer or farther from the micro-lens array will result in those objects appearing with greater or smaller size in pixels in the various micro-images. Thus, for one of these two objects to appear sharp in the generated 2D view image, different patch sizes would have to be selected (with larger and smaller sizes, respectively) from each micro-image. This means that it is possible to control the plane of focus in the generated 2D view image (i.e., which objects will appear in sharp focus) by choosing a suitable patch size. An important issue with these algorithms is that the resolution of the output 2D view depends on the selected plane of focus and, thus, on the patch size. Therefore, by varying the patch size, different content will be generated for the first two hierarchical layers presented in Figure 1;
- ii) *Position of the patch*: By varying the relative position of the patch in the micro-image, it is possible to generate 2D views with different horizontal and vertical viewing angles (i.e., different scene perspectives). Since the 3D multiview content usually represents perspectives with different horizontal angles of projection, the 3D multiview content (for the First Enhancement Layer) is generated here by varying only the horizontal position of the patch.

For the *Basic Rendering* algorithm, the input is a 3D holoscopic image with an $N_x \times N_y$ array of micro-images, where each micro-image has a resolution of $MI_x \times MI_y$. In this 3D holoscopic image, a patch of $PS \times PS$ pixels is extracted from each micro-image. These patches are then stitched together, as illustrated in Figure 2b. As a result, the output is a 2D view image with resolution of $PS \cdot N_x \times PS \cdot N_y$.

In the *Weighted Blending Algorithm*, since each micro-image captures overlapping areas of the scene, the weighted blending consists in averaging together all these overlapping regions across different micro-images, weighting differently the overlapping pixels. As can be seen in Figure 2c, each micro-image is overlapped with a shift of PS pixels

(corresponding to the patch size) to its neighboring micro-images in both horizontal and vertical directions. Then the pixels in the same spatial position across various micro-images are averaged by using a specific weighting function. Similarly to the basic algorithm, the extracted 2D view image has $PS \cdot N_x \times PS \cdot N_y$ resolution.

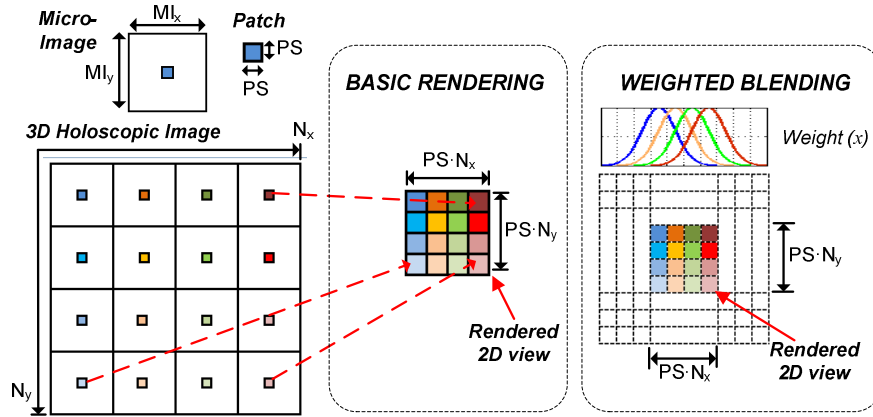


Figure 2 Generating a 2D view image: (a) 3D holoscopic input image; (b) Basic Rendering algorithm [1]; and (c) Weighted Blending algorithm [1].

3. INTER-LAYER PREDICTION ON 3D HOLOSCOPIC SCALABLE CODING

It was shown in [6] that higher coding efficiency can be achieved by exploring the existing redundancy between the *First Enhancement Layer* (multiview content) and the *Second Enhancement Layer* (3D holoscopic content), through an inter-layer prediction scheme. This prediction scheme builds an inter-layer (IL) reference picture which is then used to predict a 3D holoscopic image being coded. To build an IL reference, the following pieces of information are need:

- *Set of 2D views*: The set of reconstructed 2D views from the previous coding layers are obtained by decoding the bitstream generated in the lower layers;
- *Acquisition parameters*: These parameters comprise information from the 3D holoscopic capturing process (such as the resolution of micro-images and of the micro-lens array) and also information from the 2D view generation process (i.e., size and position of the patch, as explained in Section 2). This information has to be conveyed along with the bitstream to be available at the decoding side.

Therefore, two steps are distinguished when generating an IL reference: *Patch Remapping*, and *Micro-image Refilling*.

3.1 Patch Remapping

This first step is an inverse process of the *Basic Rendering* algorithm, presented in Section 2, i.e., it corresponds to an inverse mapping (referred to as remapping) of the patches from all rendered and reconstructed 2D views to the 3D holoscopic image.

For this step, the needed input information is:

- i) One or more reconstructed 2D views;
- ii) Patch size used to generate the 2D views;
- iii) Micro-image resolution;
- iv) Relative position of the patch in the micro-image used to take each different 2D view.

Then, each 2D view can be subdivided in patches and each patch can be mapped to its original position in the 3D holoscopic image, as illustrated in Figure 3. A template of the 3D holoscopic image assembles all patches and the output of this step is a sparse 3D holoscopic image (see Figure 3).

3.2 Micro-Image Refilling

In the *Micro-Image Refilling* step, the significant cross-correlation existing between neighboring micro-images is emulated so as to fill the holes in the sparse 3D holoscopic image (built in the *Patch Remapping* step) as much as possible. As such, the input for this step is a sparse 3D holoscopic image generated by the *Patch Remapping* step.

For each micro-image in the sparse 3D holoscopic image, an available set of pixels, such as the patch with size $PS \times PS$, is copied to a suitable position in a neighboring micro-image. Since there is no information about the disparity between objects in neighboring micro-images, this position is defined in a patch-based manner, and is given by the position of the patch being copied, shifted by $i \cdot PS$. The variable i corresponds to the relative position of the current micro-image (where the patch is copied from) to the neighboring micro-image (where the patch is pasted on). Since this copy is performed in both horizontal and vertical directions, i is defined in \mathbb{Z}^2 . Additionally, the number of neighboring micro-images where the patch will be copied to depends on the size of the micro-image and the size of the patch.

An illustrative example of this process is shown in Figure 4 for only three neighboring micro-images. The output of this step is the Inter-Layer image prediction (Figure 4). It should be noticed, that if the patches do not fill all the horizontal positions within the micro-image width, it will not possible to estimate this information at the border of the IL picture prediction, for instance, at the first column of micro-images shown in Figure 4.

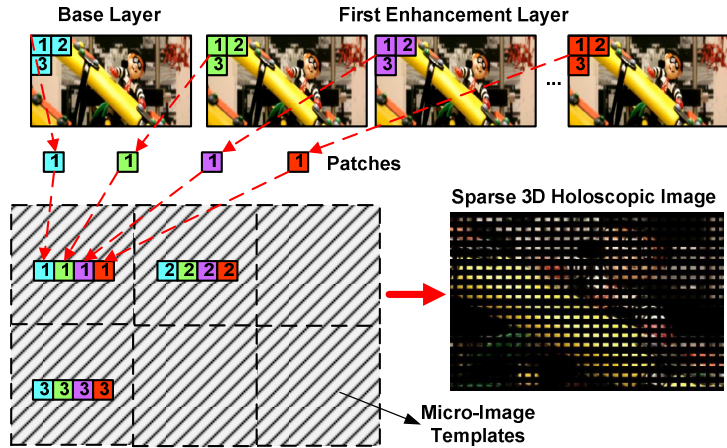


Figure 3 The *Patch Remapping* step to generate a sparse 3D holoscopic image prediction

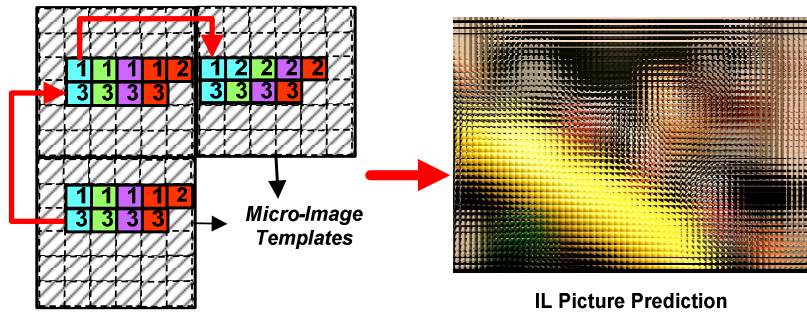


Figure 4 The *Micro-image Refilling* step to generate a 3D holoscopic prediction image

4. MITIGATION OF PACKET LOSS IMPACT ON SCALABLE 3D HOLOSCOPIC CODING

Guaranteeing successful 3D holoscopic video transmission in presence of channel errors is a challenging issue, that requires reliable error resilience mechanisms for combating the transmission errors and mitigating their impact in the user quality perception.

State-of-the-art error resilience techniques for 2D and 3D multiview video can be typically categorizes in three main groups [11]: i) error resilient encoding techniques, which are introduced into the video encoding process to make the bitstream more robust to errors; ii) error concealment techniques, which are employed at the decoding process to conceal the effect of errors; and, iii) those that require interactions between encoder and decoder to adaptively consider the network characteristics in terms of information loss.

Since there is a lack of error resilience techniques specific for 3D holoscopic content, a simple error concealment technique is proposed to estimate the lost data making use of the inherent correlation existing in the 3D holoscopic

content. In this section, a discussion about some of the relevant factors which affects the inter-layer prediction accuracy is firstly presented (in Section 4.1) and, then, the proposed error concealment method is defined (in Section 4.2).

4.1 Relevant Factors for the Inter-Layer Prediction Accuracy

Besides the aforementioned advantages of using a 3D holoscopic imaging system, it is important to notice that for representing the full light field in this type of content there is a massive increase in the amount of information that need to be captured, encoded and transmitted when compared to legacy technologies. As opposing the MVC approach where each enhancement layer represents an additional 2D view image, there is a considerable jump in the coding information amount between *First* and *Second Enhancement Layers* of the proposed scalable coding architecture.

To illustrate the relation between amounts of information in the lower hierarchical layers and the *Second Enhancement Layer*, consider one frame from the 3D holoscopic test image *Plane and Toy* (frame 123, in Figure 6a), with resolution of 1920×1088 and micro-image resolution of around 28×28 pixels. From this 3D holoscopic content, 9 views are generated for the first two scalability layers— one for the *Base Layer* and eight for the *First Enhancement Layer*. These views are generated using the *Basic Rendering* algorithm with patch size of 4×4 and varying the position of the patch in relation to the center of the micro-image in $\{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$ pixels. Notice that, from this set of patch positions, adjacent patches contain overlapping areas of the micro-image. Consequently, approximately 12% of the information inside each micro-image is used to build these nine 2D views and the remainder data is discarded. The nine views are then coded independently with the HEVC using the “Intra, main” configuration [12].

Afterwards, the nine coded and reconstructed 2D views are processed to build an IL reference. In the *Patch Remapping* step, since there are overlapping areas between adjacent patches, this redundant information is used to refine the pixel values. The resulting sparse 3D holoscopic image is shown in Figure 6b by the enlargement, to illustrate the amount of information that need to be estimated in the *Micro-Image Refilling* process. After this, by applying the *Micro-Image* refilling process, the IL picture prediction is completed, as shown Figure 6c. This IL picture prediction is then used as a new reference picture to efficiently encode the 3D holoscopic content in the *Second Enhancement Layer*, as previously proposed by the authors in [6].

Finally, in Figure 5, the used bitrate for encoding all the nine 2D view images independently is compared to bitrate used to encode the 3D holoscopic content with the scalable coding scheme for four different quantization parameter (QP) values. From this, it can be seen that the *Base Layer* and the *First Enhancement Layer* represent only a small percentage of the scalable bitstream (in this case, about 16% of the scalable bitstream). Therefore, it is expected that losses in the lower hierarchical layers will considerably affect the accuracy of the built IL picture prediction and, consequently, degrade the performance of the proposed scalable coding scheme.

Moreover, it should be also noticed that, as was shown in [6], the performance of the inter-layer prediction scheme is improved when increasing the patches sizes. This fact is related again to the amount of data from a 3D holoscopic content that is discarded when generating a 2D view image and that need to be estimated in the *Micro-Image Refilling* process. As the amount of discarded information is a consequence of the chosen patch size and number of views, this means that the parameters which are freely chosen when generating the content in the lower hierarchical layers will also affect the accuracy of the build IL picture prediction.

Considering the *Basic Rendering* and *Weighted Blending* algorithms, these parameters are:

- *Patch Size* – during the creative post-production process, a proper patch size will be selected and will be limited to the used optical depth of field. As mentioned earlier, the quality of the IL picture prediction will improve as relative larger patch sizes are used.
- *Number and position of 2D views* – the choice of number of views and their corresponding positions is based on the type of display that will be used. In this case, as the number of 2D view images increases, less information from the 3D holoscopic content will be discarded and, consequently, the quality of the IL prediction may improve. However, if these 2D views are generated by overlapping patches positions, the amount of relevant information to build the IL prediction picture is smaller, and its performance may decrease.

In other words, there is a large degree of freedom when defining how the 3D holoscopic content will be presented. Therefore, the error resilience problem need to be analyzed taking into account the parameters that control the generation of content for the lower hierarchical layers, since the quality of the inter-layer prediction is also dependent on them and may also affect the effectiveness of a resilience error technique.

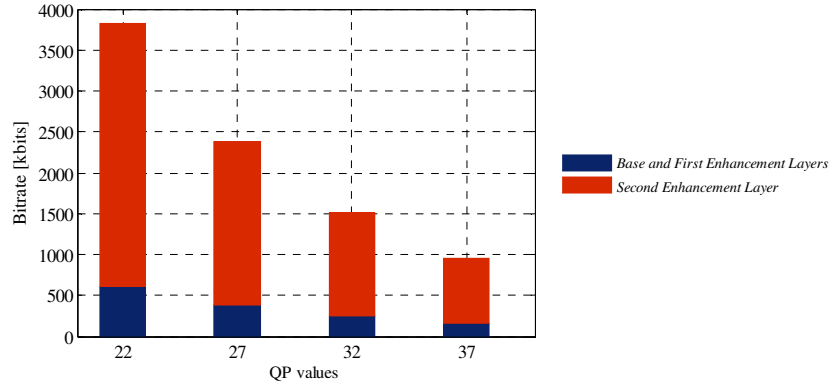


Figure 5 Relation between the amount of data in the bitstream for the Base Layer and *First Enhancement Layer*, compared to the *Second Enhancement Layer*.

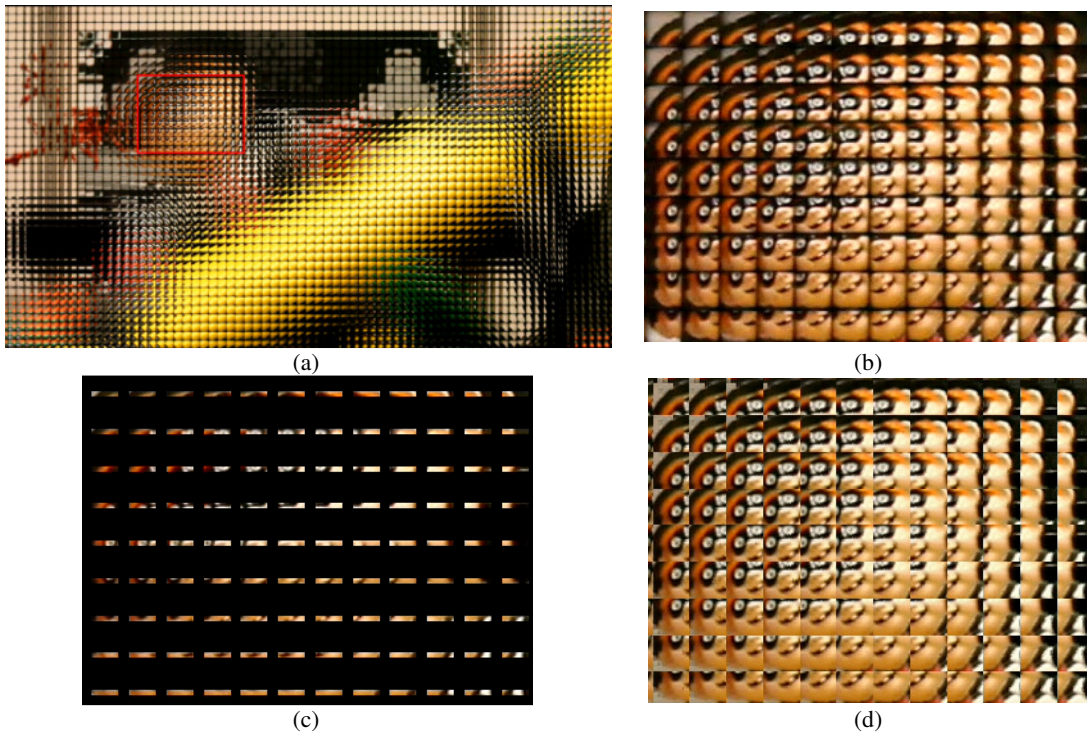


Figure 6 3D Holoscopic Image and corresponding IL picture prediction: (a) 3D holoscopic image *Plane and Toy* (frame 123); (b) Magnified section of 336×246 pixels from original image; (c) Magnified section of 336×246 pixels from sparse IL prediction picture; and (d) Magnified section of 336×246 pixels from full IL prediction picture.

4.2 Proposed Error Concealment Algorithm

Typically, an error concealment algorithm makes use of spatial, temporal, and spectral redundancy of the content to estimate the missing data and mask the effect of channel errors at the decoder.

Although the conventional error concealment tools for the lower layers in the hierarchical scalable architecture can be applied to the *Second Enhancement Layer*, these methods do not consider the inter-layer correlation between the multiview and 3D holoscopic content, and neither the inherent spatial correlation of the 3D holoscopic content.

When generating the IL picture prediction, the *Micro-Image Refilling* process is already able to estimate non-existing data to fill the holes in the sparse 3D holoscopic image, by making use of the significant cross-correlation existing between neighboring micro-images. Therefore, considering that a 2D view image is lost, the only difference when

building the IL picture prediction is that there will be more holes to be fulfilled in the *Micro-Image Refilling* process. This means that, it is possible to simply derive the error concealment algorithm from the IL prediction method.

Therefore, upon the detection of a lost picture, the following steps are considered by the proposed error concealment algorithm to build the IL picture prediction:

1st) The *Patch Remapping* process is employed considering only the set of 2D view images that are available (without loss). To illustrate the consequence of a lost 2D view in this step, five non-overlapping patch positions ($\{-8, -4, 0, 4, 8\}$) are used to generate five corresponding 2D view images from the 3D holoscopic image *Plane and Toy* (frame 123). Then, considering that the central 2D view image (with patch position “0”) is not available at the decoder side, the sparse IL picture prediction will contain extra holes where the patches of the lost 2D view were supposed to be placed, as illustrated in Figure 7a. In this example, a magnified section of the sparse IL prediction picture is shown, where the set of non-available pixel positions are represented in green.

2nd) The *Micro-Image Refilling* algorithm is able to estimate most of the holes by using information from available patch positions, including the set of lost patches in the position “0”. This is illustrated in one of the steps of the algorithm (for the first 2D view) in Figure 7b. Finally, it is possible to re-call the algorithm also for the lost patch position to fulfill the IL prediction picture, as shown in Figure 7c.

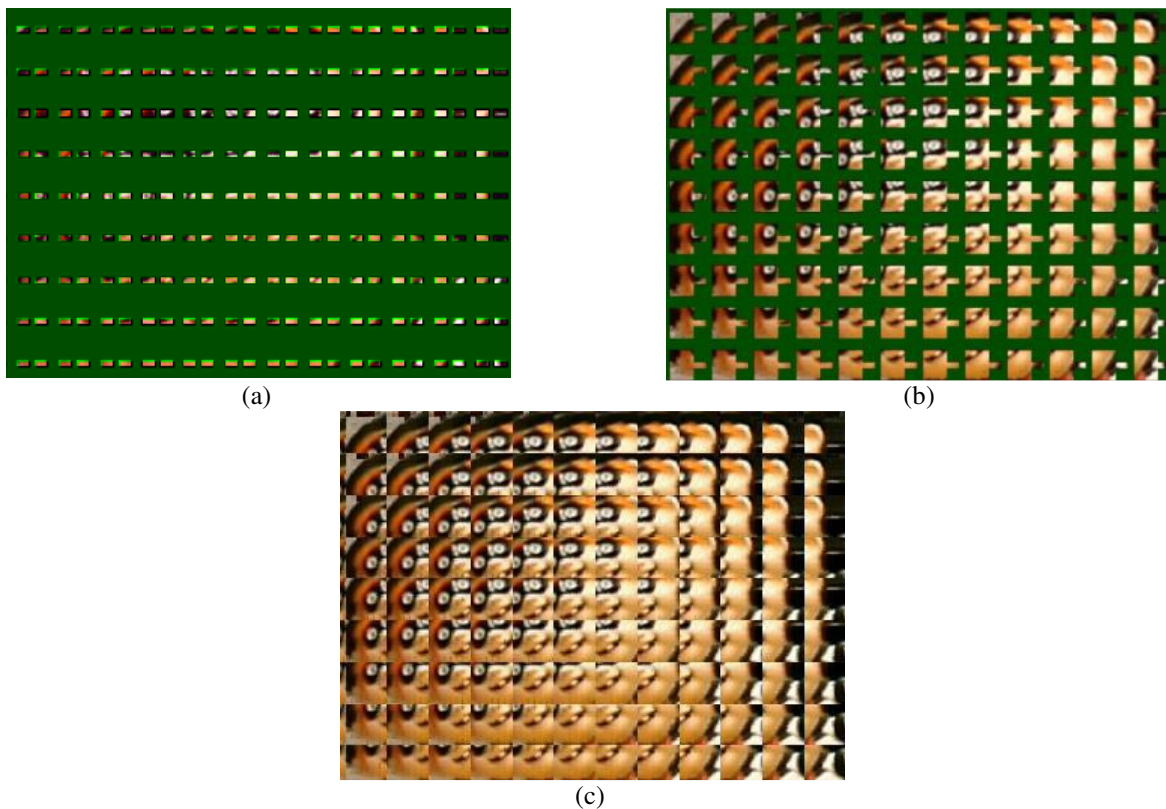


Figure 7 Some steps of the used error concealment algorithm to build the IL reference when one 2D view image is lost: (a) The *Patch Remapping* for the available 2D view images; (b) One of the iterations of the *Micro-Image Refilling* to illustrate the recovery of the lost patches; and (c) The built IL picture prediction.

5. EXPERIMENTAL RESULTS

In this section, a study of how packet losses affects the accuracy of the inter-layer prediction in the *Second Enhancement Layer* is presented, highlighting the influence of chosen rendering algorithm, patch size, and number and position of the generated 2D view images on this. For this analysis, the considered test conditions are presented in Section 5.1 and, then, the results analysis is shown in section 5.2.

5.1 Test Conditions

In order to properly analyze the influence of packet losses on the accuracy of the Inter-Layer prediction, four 3D holoscopic test images were used so as to achieve a set of representative results:

- i) Two frames of the sequence *Plane and Toy* (frames 23 and 123, as shown in Figure 8a and b, with resolution of 1920×1088, micro-image resolution of 27.75×27.75. In these frames, the main object of the scene (the toy on the plane) appears in different plane of focus.
- ii) *Robot 3D*, with resolution of 1920×1088, micro-image resolution of 27.5×27.5, as illustrated in Figure 8c;
- iii) *Laura*, with resolution of 7240×5432, micro-image resolution of 74.75×74.75 as shown in Figure 8d (available in [13]).

To generate the content for the first two scalability layers, the four test images were processed with both algorithms presented in Section 2 (*Basic Rendering* and *Weighted Blending*). In this process, nine 2D view images were generated – one for the *Base Layer* and eight for the *First Enhancement Layer*. Since the resolution of the micro-images varies from one image to another, the patch positions to generated 2D view images were chosen so as to have nine regularly spaced views within the micro-image limits.

Three different patch sizes were chosen for each test image, which correspond to cases where adjacent patches are taken with and without overlapping areas. Additionally, one of these patch sizes represents the case where the main object of the scene is in focus. Due to the small size of micro-images in *Plane and Toy* and *Robot 3D* images, an additional set of patch positions needed to be considered so as to have the case where the patches are taken without overlap areas. In this case, five regularly spaced 2D view images were generated.

Therefore, the chosen patch sizes and positions for each tested 3D holoscopic image are summarized in Table 1.

Table 1 Tested Conditions: Patch Positions and Patch Sizes for each tested 3D holoscopic image.

Tested Image	Patch Positions	Patch Sizes
<i>Plane and Toy</i> (frame 23)	9 views $\{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$	4 (in focus), 9 and 11
	5 views $\{-8, -4, 0, 4, 8\}$	
<i>Plane and Toy</i> (frame 123)	9 views $\{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$	4, 9 (in focus) and 11
	5 views $\{-8, -4, 0, 4, 8\}$	
<i>Robot 3D</i>	9 views $\{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$	4 (in focus), 9 and 11
	5 views $\{-8, -4, 0, 4, 8\}$	
<i>Laura</i>	9 views $\{-28, -21, -14, -7, 0, 7, 14, 21, 28\}$	7, 10 (in focus) and 14

It should be noticed that, due to the large number of possible combinations of test conditions (number of views, patch size and patch positions) and since this paper is mainly focus on analyzing the influence of these parameters on the performance of the error concealment algorithm, it will not yet cover an extensive set of network conditions, which will be, however, considered in future work. To simulate the network conditions, it is considered that an entire 2D view image is coded into only one packet. Hence, loss of a packet implies that the entire 2D view image must be recovered by the error concealment algorithm. Three different packet loss conditions were considered, where one, two and three packets are lost. Additionally, packet losses were assumed independent and identically distributed for all 2D view images. For this, it is considered a case where the two lower layers are independently encoded, since an enhancement layer would not be decodable if the 2D view image in the base layer was lost.

The results are presented in terms of the average Mean Squared Error (MSE) (for all the combination of lost 2D view images) of the IL picture prediction built by the error concealment algorithm, compared with the IL prediction picture when there is no packet loss. Alternatively, the average MSE is also shown discarding the cases where the first or the last pictures are lost, since when this happens, portion of the information cannot be recovered by the *Micro-Image Refilling* algorithm in the border of the IL picture reference.

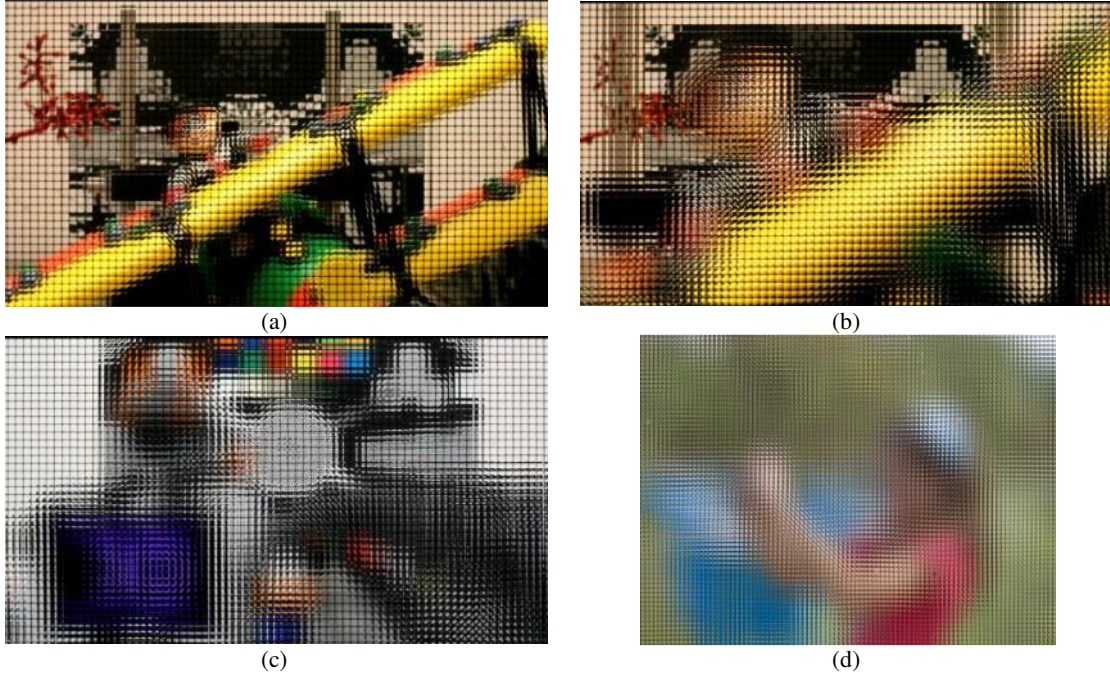


Figure 8 Test images for the scalable 3D holographic codec evaluation: (a) *Plane and Toy*; (b) *Scene1*; and (c) *Laura*

5.2 Results Analysis

The experimental results for each tested 3D holographic image can be seen in Figures 9 to 15. In each Figure (9 to 15), these results are split in different charts for each used rendering algorithm (*Basic Rendering* and *Weight Blending* algorithms) and for each patch size. Finally, each chart shows the corresponding average MSE value for all the possible combinations of lost 2D views (referred to as *All Views*) as well as the average MSE value when discarding the cases where the first or the last pictures are lost (referred to as *Without Border Views*). Additionally, the maximum and minimum MSE values in each case are also presented by the error bars.

Based on these charts, the following conclusions can be drawn in terms of:

- *Number of lost 2D view images*: This analysis compares how the accuracy of the built IL picture prediction varies when different numbers of 2D view images are lost. As expected, in all test conditions, the accuracy of the inter-layer prediction degrades as the number of lost views increases. For instance, considering the tested image *Laura* when using the larger patch size (14) to generate the 9 views with the *Basic Rendering algorithm* (in Figure 15a), the average MSE value goes from 65.23, when only one view is lost, to 253.75, when three views are lost. Moreover, it can be seen that the influence of lost views are stronger when the first or the last 2D views are lost. For example, for the same abovementioned test condition, the corresponding average MSE values for the *Without Border Views* case are considerably smaller (respectively, 5.9 and 61.63 when one or three views are lost).
- *Different patch sizes*: This analysis compares the results when using different patch sizes, for each tested image with the same patch positions and rendering algorithms. Surprisingly, for all results, the patch size corresponding to the case where the main object is in focus was shown to be the less affected by lost 2D views, even when it is the smaller one. For instance, consider the results shown in Figure 11 for *Plane and Toy* (frame 123), where 9 views were generated with the *Basic Rendering algorithm*. The patch size 4, where the main object is in focus, presented smaller average MSE values than the presented when using larger patch sizes. However, it is known that in this case (patch size 4), more information from the 3D holographic image was discarded from the original 3D holographic image (when generating the 2D views) and need to be estimated in the *Micro-Image Refilling* process. From this, it can be concluded that, in a sequence where there is interest in varying the patch sizes from one frame to another (e.g., the *Plane and Toy* sequence), the impact of losses will be considerably lower, since the main object is maintained in focus (which is, most of the times, the case).

- *Different rendering algorithms:* This analysis compares the results when using one of the rendering algorithms, *Basic Rendering* (in Figures 9a to 15a) and *Weighted Blending* (in Figures 9b to 15b), for each tested image and test conditions shown in Table 1. From this, it can be seen that the accuracy of the inter-layer prediction using the *Weighted Blending* algorithm are generally better than using the *Basic Rendering* algorithm when one or more views are lost. This can be explained by the high level of blur which is introduced by the weighted average in the *Weighted Blending* algorithm. Hence, the differences in these blurred images will be less evident than differences in images generated by the *Basic Rendering* algorithm.
- *Different number of 2D view images in the lower layers:* This analysis compares the results when different numbers of 2D views are generated to the lower layers, using the same patch sizes and rendering algorithms. For this, the results using 5 and 9 views for Plane and Toy (frame 23) (in Figure 9 and 10), Plane and Toy (frame 123) (in Figure 11 and 12), and Robot (in Figure 13 and 14) are compared. As expected, by using less 2D view images in the lower layer, a loss of 2D view images will affect more drastically the accuracy of the built IL picture prediction. This can be understand since, when considering less views: i) more information from the original 3D holoscopic image is discarded to generate the 2D view images; and ii) a loss of a 2D view image represent a higher packet loss rate.

It is important to notice that, although for these tests it was considered that an entire 2D view image is coded into only one packet, the *Patch Remapping* and *Micro-Image Refilling* processes could easily be adapted to the case where the lost packets represents slices of 2D view images. Moreover, from the presented analysis, it was shown that, although the parameters of the scalable coding architecture somehow interfere on the performance of the error concealment algorithm, in some cases, the used error concealment algorithm is able to recover the IL picture prediction with negligible MSE value (e.g. when the *Weighted Blending* algorithm is used). However, it is also important to consider cases where the losses happen in the *Second Enhancement Layer* of the proposed scalable coding solution, since the information in this layer represents the largest percentage of the scalable bistream (as shown in Section 4.1). Therefore, these cases will be considered in future work.

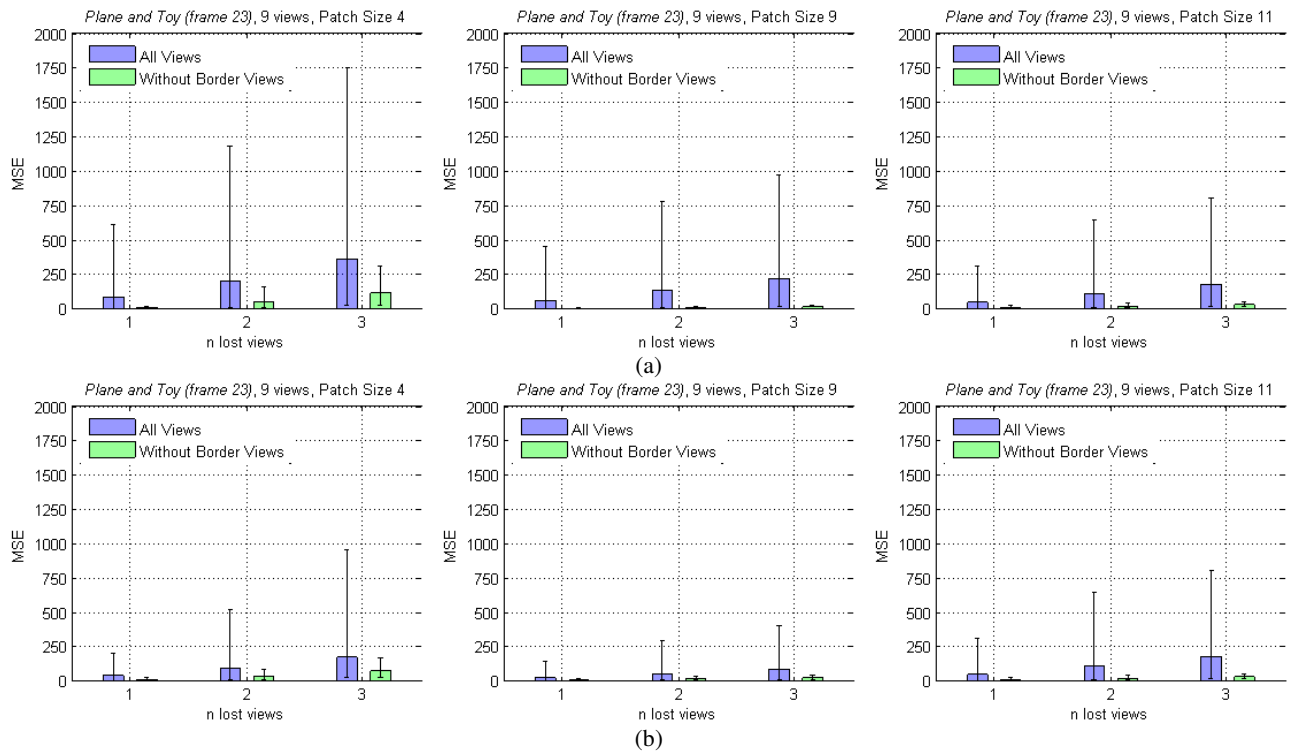


Figure 9 Comparison between qualities of the IL reference when there are lost views. In this case, 9 views were generated with 3 different patches from the tested holoscopic image *Plane and Toy* (frame 23) using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

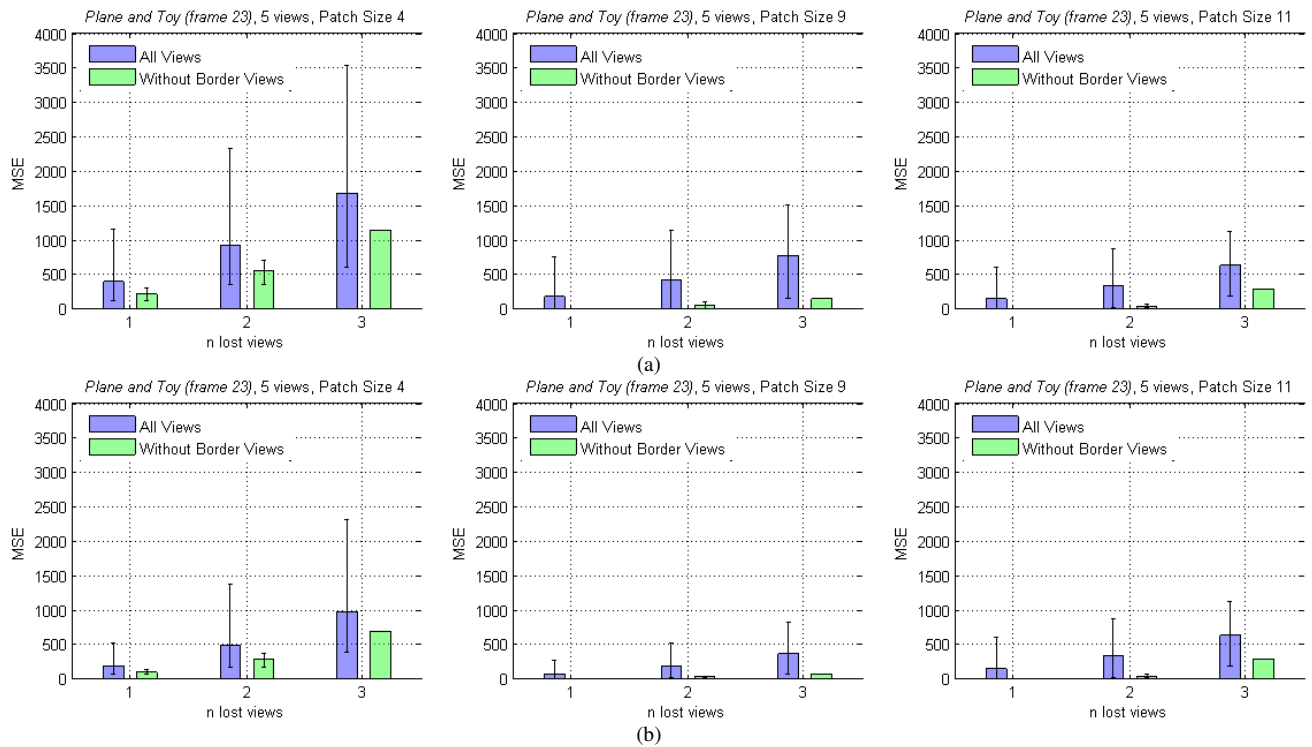


Figure 10 Comparison between qualities of the IL reference when there are lost views. In this case, 5 views were generated with 3 different patches from the tested holoscopic image *Plane and Toy* (frame 23) using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

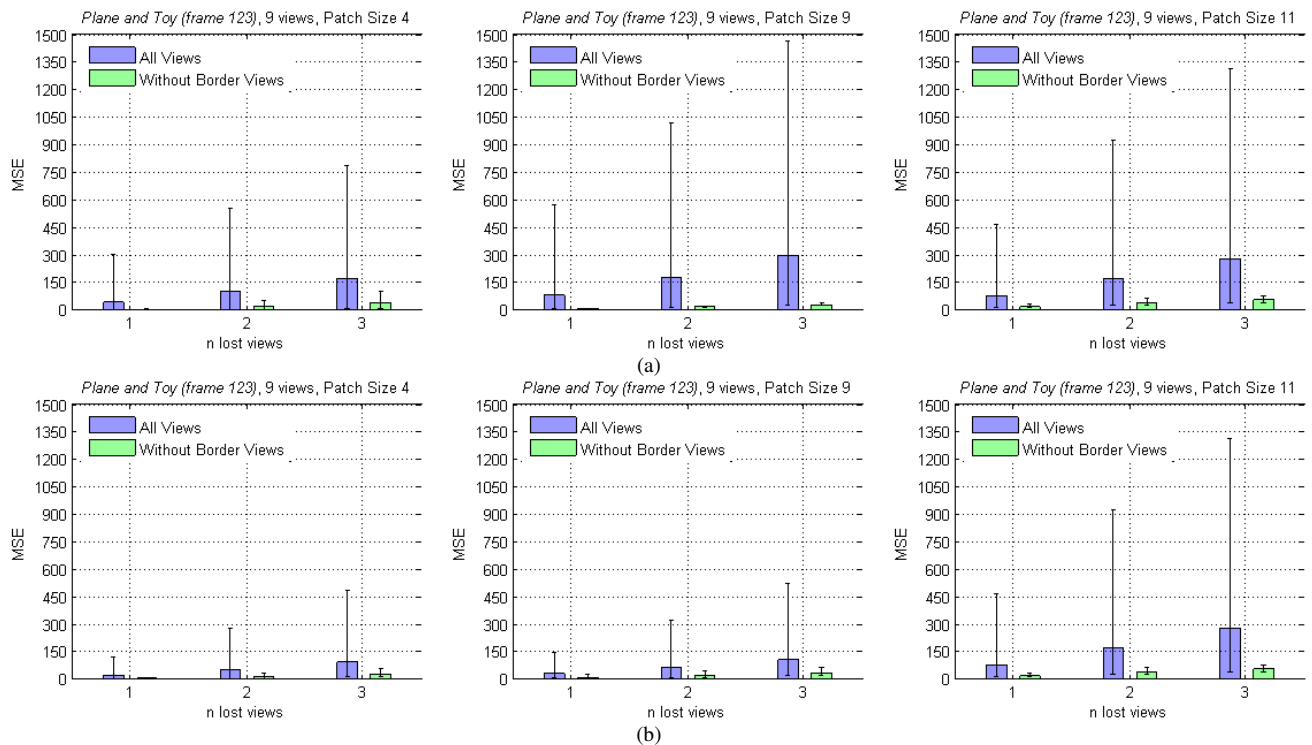


Figure 11 Comparison between qualities of the IL reference when there are lost views. In this case, 9 views were generated with 3 different patches from the tested holoscopic image *Plane and Toy* (frame 123) using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

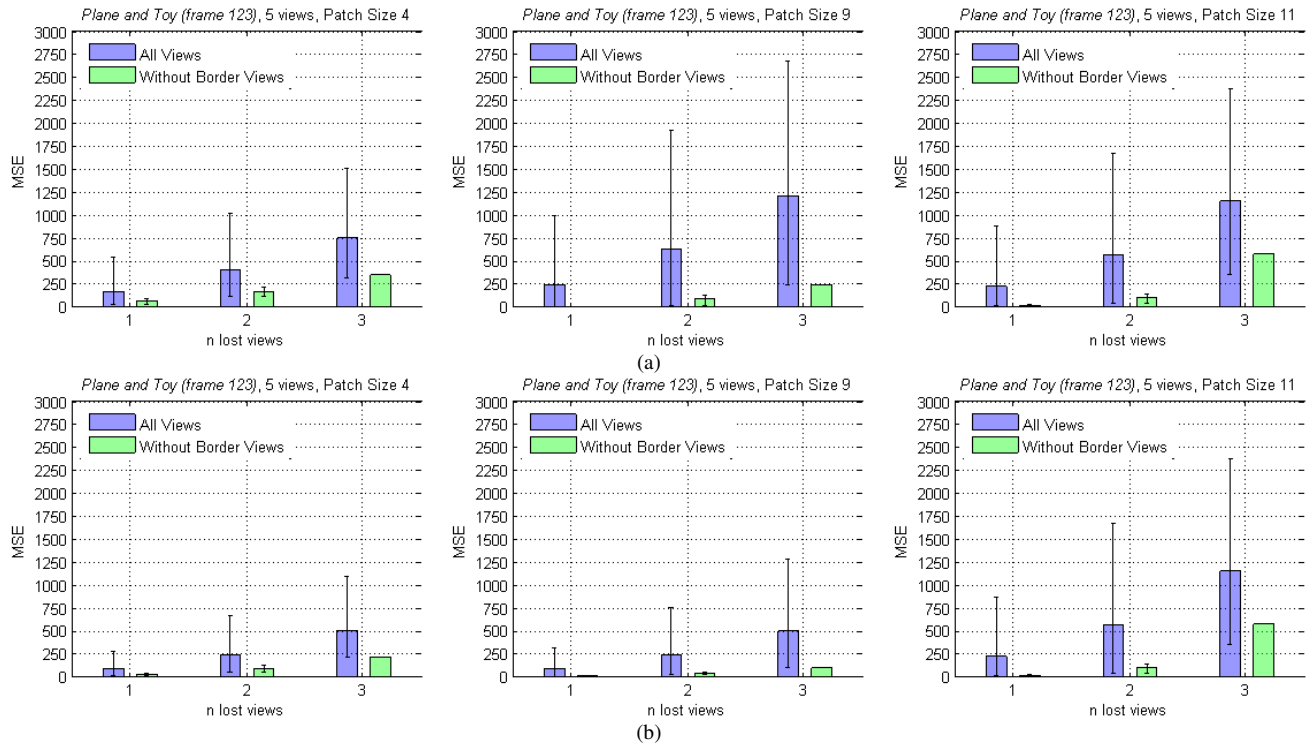


Figure 12 Comparison between qualities of the IL reference when there are lost views. In this case, 5 views were generated with 3 different patches from the tested holoscopic image *Plane and Toy* (frame 123) using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

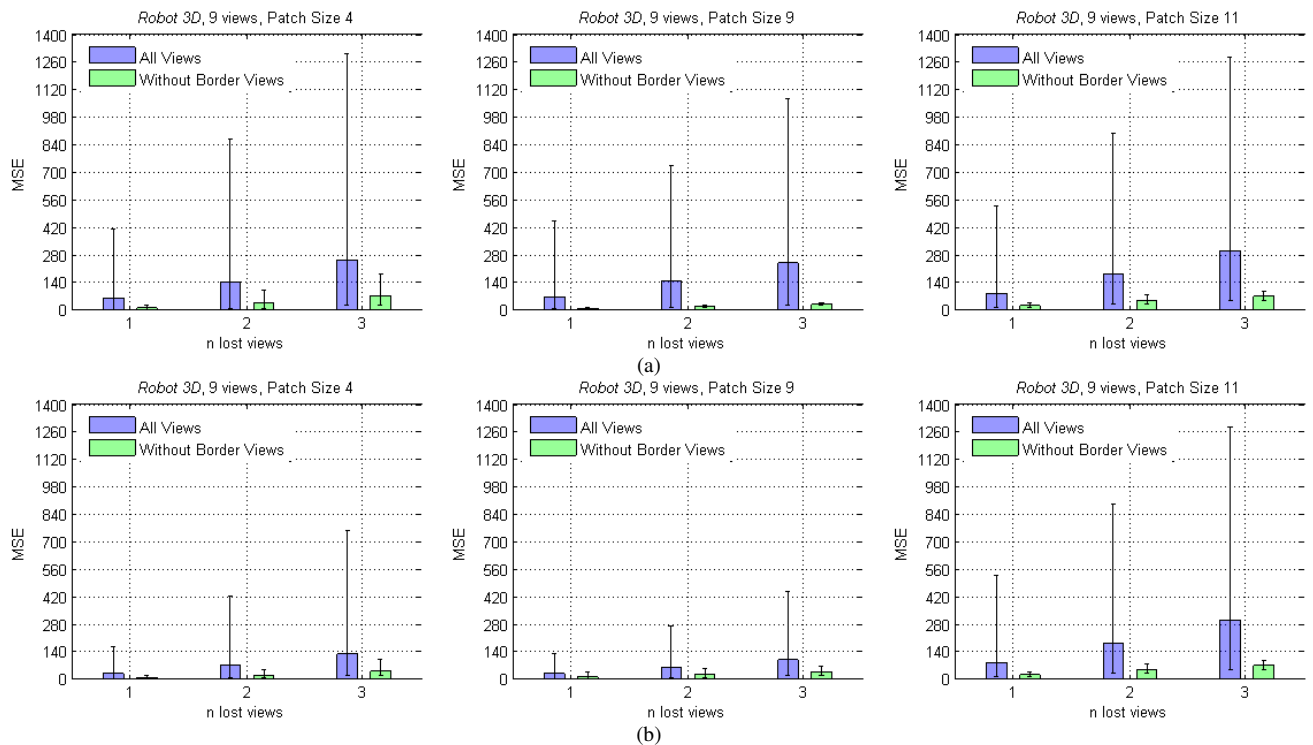


Figure 13 Comparison between qualities of the IL reference when there are lost views. In this case, 9 views were generated with 3 different patches from the tested holoscopic image *Robot 3D* using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

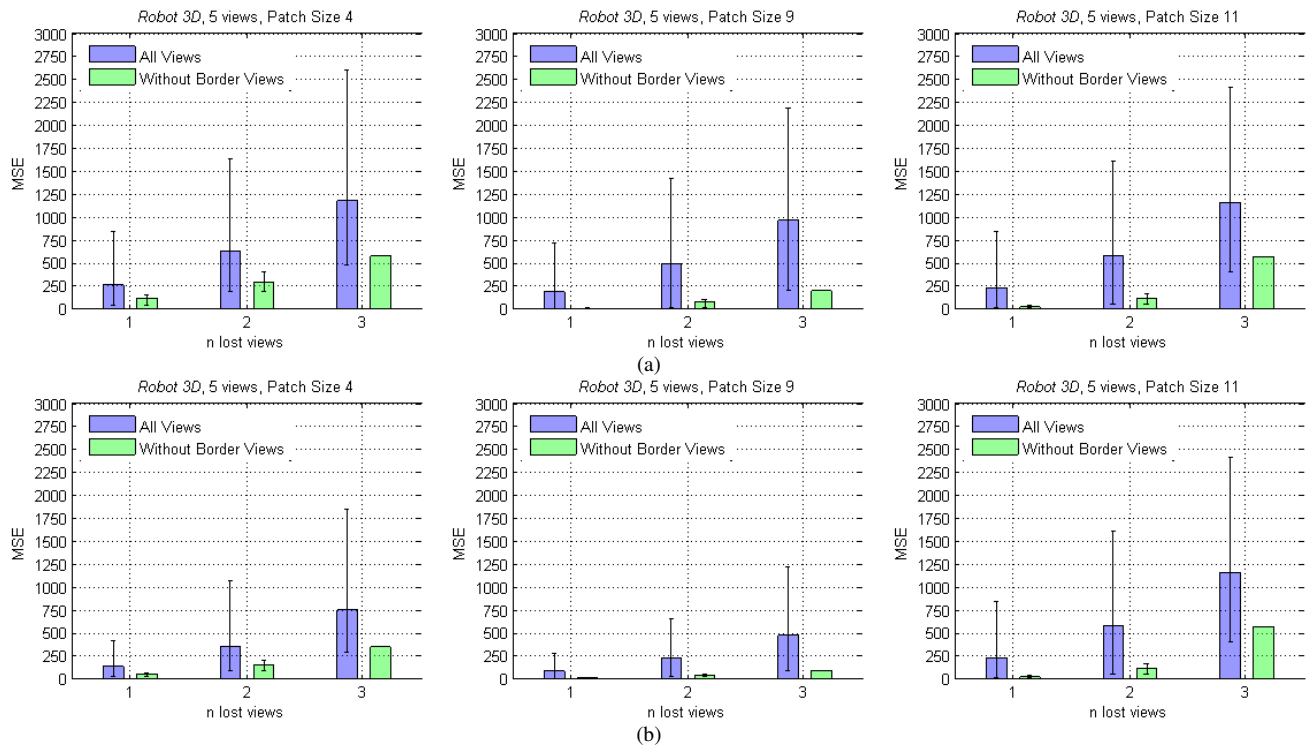


Figure 14 Comparison between qualities of the IL reference when there are lost views. In this case, 5 views were generated with 3 different patches from the tested holoscopic image *Robot 3D* using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

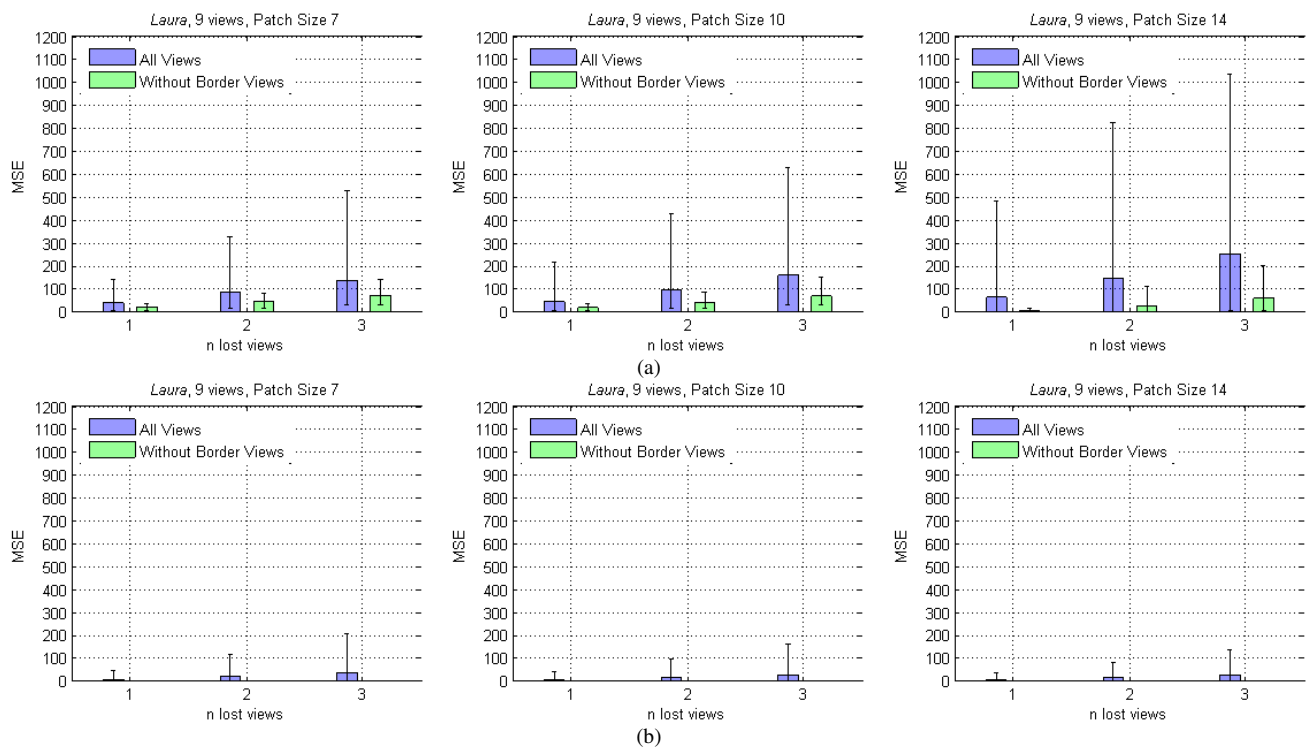


Figure 15 Comparison between qualities of the IL reference when there are lost views. In this case, 9 views were generated with 3 different patches from the tested holoscopic image *Laura* using: (a) *Basic Rendering* algorithm; and (b) *Weighted Blending* algorithm.

6. FINAL REMARKS

This paper proposed to start the discussion about error resilience techniques for 3D holoscopic content and presents a study of the influence of packet losses in display scalable 3D holoscopic content coding. For this, an error concealment algorithm was proposed to estimate the lost data, which was derived from the inter-layer prediction scheme previously proposed by the authors. From the presented study, it could be seen that although the parameters of the scalable coding architecture somehow interfere on the performance of the error concealment algorithm, in some cases, it is possible to recover the inter-layer prediction with negligible differences compared to the prediction when there are no losses. However, it is also important to consider cases where the losses happen in the *Second Enhancement Layer* of the proposed scalable coding solution. Therefore, future work includes proposal of error resilience techniques for dealing with transmission errors in this layer.

7. ACKNOWLEDGEMENTS

This work was supported by the European Commission under the FP7 project 3D VIVANT (Live Immerse Video-Audio Interactive Multimedia), and by the FCT (Fundação para a Ciência e a Tecnologia, Portugal) under the SFRH/BD/79480/2011 and PEst-OE/EEI/LA0008/2013 projects.

REFERENCES

- [1] T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging* 9(2), 021106–021106 (April 2010).
- [2] A. Aggoun et al., "Immersive 3D Holoscopic Video System," *MultiMedia*, IEEE 20(1), 28-37 (Jan.-Mar. 2013).
- [3] R. Ng, M. Levoy, M. Bredif, G., Horowitz, M. Duval, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," *Stanford Computer Science*, Tech. Rep. CSTR 2005-02 2005.
- [4] Raytrix GmbH. "Ratrix Lightfield Camera", <http://www.raytrix.de/index.php/Technology.html> (2012).
- [5] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018," Cisco Systems, Inc., San Diego, White Paper (2014).
- [6] C. Conti, P. Nunes, and L.D. Soares, "Inter-Layer Prediction Scheme for Scalable 3-D Holoscopic Video Coding," *IEEE Signal Processing Letter* 20(8), 819-822 (Aug. 2013).
- [7] H. Schwarz et al., "Extension of High Efficiency Video Coding (HEVC) for multiview video and depth data," in *19th IEEE International Conference on Image Processing (ICIP 2012)*, Orlando, USA, 205–208 (2012).
- [8] C. Conti, P. Nunes, and L. D. Soares, "Using Self-similarity Compensation for Improving Inter-Layer Prediction in Scalable 3D Holoscopic Video Coding," in *Proc. SPIE 8856, Applications of Digital Image Processing XXXVI*, San Diego, USA, 88561K (Sep. 2013).
- [9] M. Harris, "Focusing on everything," *Spectrum*, IEEE 49(5), 44-50 (May 2012).
- [10] T. Georgiev and A. Lumsdaine, "Rich image capture with plenoptic cameras," in *IEEE ICCP 2010*, 1-8 (Mar. 2010).
- [11] Y. Wang, S. Wenger, W. Jiantao, and A.K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, 17(4), 61-82 (Jul. 2000).
- [12] Frank Bossen, "Common HM test conditions and software reference configurations," Document JCTVCL1100 (2013).
- [13] T. Georgiev, "Todor Georgiev's Webiste", <http://www.tgeorgiev.net/> (Jan. 2013)