Study of Coding Units Depth for Depth Maps Quality Scalable Compression Using SHVC

Dorsaf Sebai[a], Faouzi Ghorbel and Sounia Messbahi

Cristal Laboratory, National School of Computer Science, Manouba University, Tunisia

Keywords: Scalable High Efficiency Video Coding, Depth Maps, Coding Units, Quality Scalability.

Abstract: Scalable High Efficiency Video Coding (SHVC) is used to adaptively encode texture images. SHVC architecture is composed of Base and Enhancement Layers (BL and EL), with an interlayer picture processing module between them. In order to ensure effective encoding, each picture is divided into a certain number of Coding Units (CUs), with different depths, composing the Coding Tree Unit (CTU). Being initially dedicated to texture images, SHVC does not provide the same efficiency when applied to depth maps. To understand the causes behind, we propose to study the SHVC CTU partitioning for depth maps. This can be a starting point to propose an efficient 3D video scalable compression. Main observations of this study show that the depth of most CUs is 2 and 3 for texture images. However, this depth is either 0 or 1 for depth maps. Moreover, CUs depths frequently change when passing from the base and enhancement layers of SHVC for the non-flat regions. This is not the case for the smooth regions that generally preserve the same CUs depths in the two SHVC layers.

1 INTRODUCTION

3D video has become increasingly popular with advances in communication, display and related areas. Today, 3D video leads to the emergence of new technologies, namely virtual, augmented and mixed realities that find their applications in several fields such as health, education, and industry. Even for Internet of Things (IoT), the future is for the 3D vision and IoT with depth to allow machines, such as autonomous cars, robots and drones, a deep perception like human beings. This is all the more true as cameras, which simultaneously capture the image and its depth, become more and more accessible to the general public thanks to their integration in smartphones.

The digital age has greatly changed the consumption of 3D video content, defining new trends and constraints that the standards of compression must face. 3D videos have in fact become accessible on many devices, such as television, computer, smartphones, IoT gadgets and by many transmission media such as the Internet, mobile, terrestrial and satellite networks. At the same time, users are increasingly demanding good quality. This is especially true with the emergence of new video formats, such as Ultra High Definition (UHD), High Dynamic Range (HDR) and High Frame Rate (HFR). Faced to these challenges, scalable compression is required to provide multiple streams of the same video that meet the heterogeneous needs of the receivers.

Being a scalable extension of the High Efficiency Video Coding (HEVC) (Sullivan et al., 2012) video compression standard, the Scalable High Efficiency Video Coding standard (SHVC) (Boyce et al., 2016) makes it possible to perform scalable encoding. SHVC is dedicated to the scalable compression of conventional 2D videos whose only component is texture images. It is not adapted to depth maps as it induces damaging visual artifacts at sharp depth discontinuities (Sebai, 2020). Further, 3D High Efficiency Video Coding (3D-HEVC) (Tech et al., 2016), is the latest standard dedicated to the compression of depth maps. But, it does not allow scalable compression of these latter. However, the need for scalable compression also persists for 3D video used by many applications, such as virtual, augmented and mixed reality. In this paper, we propose a study of the SHVC CTU partitioning for depth images in order to evaluate its efficiency in 3D context. Firstly, we aim to analyse the CTUs depth in texture images versus depth maps. Secondly, we analyze the CTU construction in Base Layer (BL) versus Enhancement Layer (EL) of SHVC.

The rest of this paper is organized as follows. In Section 2 and 3, we present the concepts required to

[a] https://orcid.org/0000-0001-7720-2741

Sebai, D., Ghorbel, F. and Messbahi, S.
Study of Coding Units Depth for Depth Maps Quality Scalable Compression Using SHVC.
DOI: 10.5220/0011706400003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 114-120
ISBN: 978-989-758-634-7; ISSN: 2184-4321

understand the study of CUs structure which is detailed in Section 4. The conclusion is drawn in Section 5.

2 MULTIVIEW VIDEO PLUS DEPTH

Multiview video plus depth (MVD) (Merkle et al., 2007) is a 3D video format which consists of more than two streams videos each including two components, texture images and their corresponding depth maps. 3D-HEVC, the latest 3D video coding standards, is developed to efficiently compress MVD data, especially depth maps. These latter are two-dimensional representation of the scene geometry. For each texture pixel, a corresponding pixel exists in the depth map. The depth pixel value is the distance between a point in 3D space, captured by the corresponding texture pixel, and the camera. Knowing the intrinsic and extrinsic parameters of the camera, this depth value makes it possible to project a pixel into 3D space. If a view synthesis is desired, the pixel is then projected onto a virtual viewing plane thereby generating a new virtual view.

As shown in Fig. 1, depth maps are represented as 2D grayscale images. This representation allows 256 depth values, ranging from 0 to 255. The intensity value 255 defines the closest scene objects to the camera. The farthest objects correspond to the gray level 0. Being different from texture images, depth maps are distinguished by their piecewise planar definition and the impact of depth discontinuities on the quality of the synthesized views. Indeed, each plane corresponds to an object of the scene where intensities of its coplanar pixels vary in a regular way; and contours, placed at the objects edges, reproduce sharp depth discontinuities between objects in foreground and background (Yea and Vetro, 2009). These discontinuities must be preserved as their compression results in a highly visible degradation of the synthesized views. However, it is not the case for smooth regions. The distortions of compression errors in these regions have a less noticeable or even limited effect.

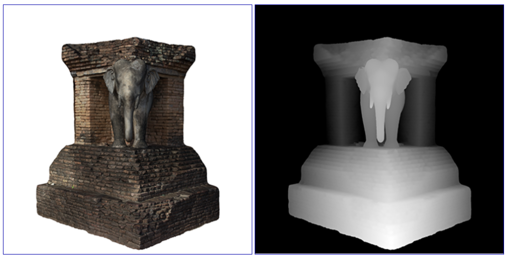


Figure 1: Example of a texture image (left) and its associated depth map (right).

3 SCALABLE VIDEO COMPRESSION

Scalability consists in extracting several versions from a stream according to the users needs, terminals capacities, and networks conditions (Tohidypour, 2016). In the particular case of images, a stream is said to be scalable when parts of this latter can be deleted, so that the resulting sub-stream forms another valid stream for a target decoder. A Base Layer (BL) is produced, and other Enhancement Layers (EL) are used to best adapt the capabilities of the receiver.

3.1 Scalable High Efficiency Video Coding

The SHVC (Boyce et al., 2016) encoder is the extension of HEVC which currently offers different types of scalability such as temporal scalability, spatial scalability and quality scalability also known as Signal to Noise Ratio (SNR) scalability. The principle of SHVC consists in encoding each image of the input video into a hierarchy of layers, starting from a BL that is gradually refined by the addition of ELs.

SHVC encodes input images in the BL bit stream using the HEVC or Advanced Video Coding (AVC) codec (Wiegand et al., 2003). The encoding of the EL is performed using HEVC while leveraging the data already encoded in the lower reference layers in order to eliminate redundancies. This Inter Layer Processing (ILP) block improves the efficiency of SHVC in terms of compression ratio. It ensures that data already encoded in the previous layers is no longer encoded in the current layer. The bit streams of the BL and EL are finally multiplexed according to the increasing order of the layers indices. On decoding, these streams are demultiplexed in order to decode each of them taking into consideration the redundancies retained by the ILP step.

3.2 Coding Tree Unit

In SHVC, a picture is divided into a quadtree, known as CTU. This latter is composed of CUs that are similar to macroblocks used in the previous AVC. Whereas macroblocks can span 4×4 to 16×16 block sizes, CTUs can process as many as 64×64 blocks. This gives it the ability to compress information more efficiently. Fig. 3 represents an illustration of the

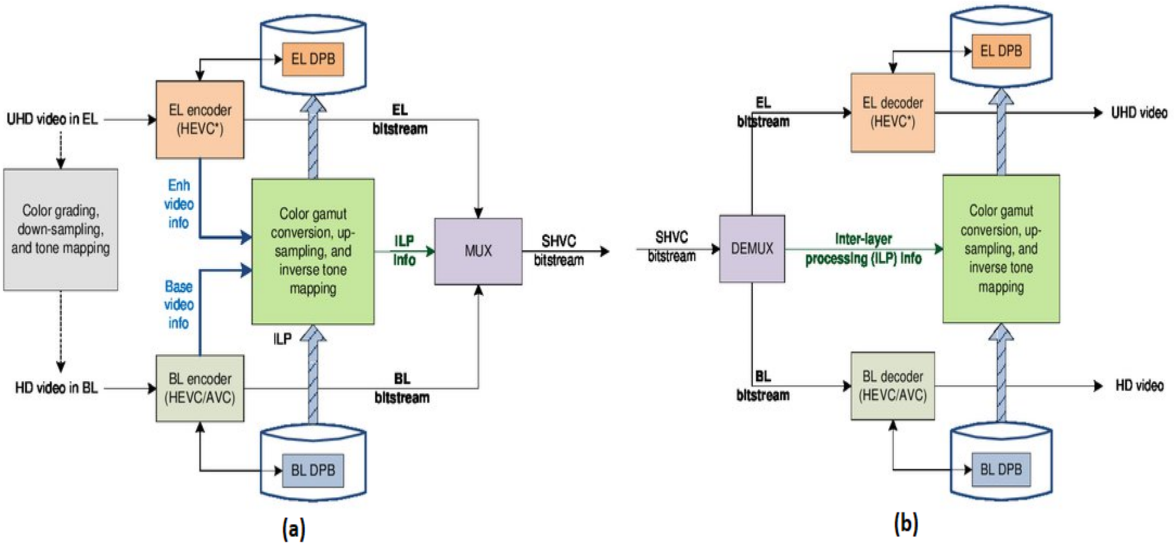


Figure 2: Architecture of SHVC: (a) SHVC encoder with two layers and (b) SHVC decoder with two layers (Gudumasu et al., 2015).

SHVC CTU structure. Each CU can be splitted into four sub-CUs due to a Rate Distortion Optimization (RDO) decision function (Sullivan and Wiegand, 1998). This ergodic process looks for all possible CUs to choose the ones with the smallest rate distortion cost. The maximum depth partitioning is set equal to 4, numbered from 0 to 3. The CTU depth (CTU Depth) 0 corresponds to the CU size of 64 × 64, CTU depth 1 to CU size of 32 × 32, CTU depth 2 to CU size of 16 × 16 and CTU depth 3 to CU size of 8 × 8. A CU will be divided into Prediction Units (PUs) that are used to carry the information related to the prediction modes. To choose the best CTU structure, the Rate Distortion Cost (RDCost) is calculated for each mode, and, at the end of the process, the partition that introduces the lowest RDCost is selected.

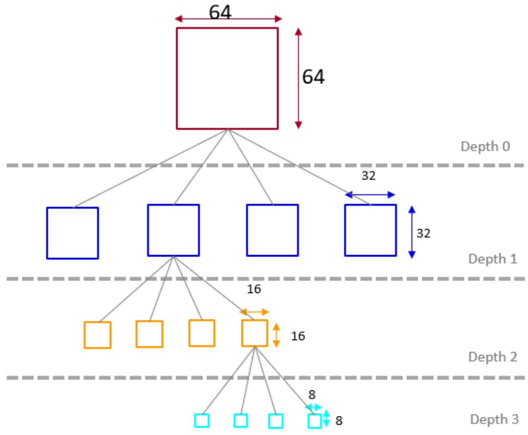


Figure 3: Illustration of the Coding Tree Unit structure in SHVC.

4 STUDY OF CODING TREE UNIT STRUCTURE

Since it was originally designed for texture images, SHVC is analyzed for depth maps. We aim specifically to study the size of a CU between the base and enhancement layers on the one hand and in the texture image and depth map on the other hand. For this syudy, We use twenty images, ten texture images and their corresponding depth maps as shown in Fig. 4. To encode these latter, the SHVC Test Model reference encoder (SHM 12.0) is used. For each of the images, two layers BL and EL are respectively coded at QP values of 26 and 22.

Figure 4: Test images: Texture images (first row) and their associated depth maps (second row).

4.1 Coding Tree Unit Structure: Base Layer vs. Enhancement Layer

In this present section, we seek to evaluate the amount of depth maps regions for which the CTU depth does not change between BL and EL. Fig. 5 and Table 1 show the percentages of CTUs that maintain the same depth when passing from BL to EL (% Regions CTU Depth BL EL). For smooth regions, we note a high correlation between base and enhancement layers. In fact, 85% of CTUs maintain their depth since object plans of depth maps are mostly encoded from the BL and no more partitioning is required for the EL. Unlike smooth regions, less than 29% of contours regions maintain the same CTU depth in EL relatively to BL. This type of regions often needs to be more splitted in EL, as the CTU depth in BL is insufficient. This observation is relevant to reduce the computational complexity of SHVC for depth maps coding. A further work can, for example, aim at proposing an algorithm that skips the CTU partitioning for smooth regions in ELs.

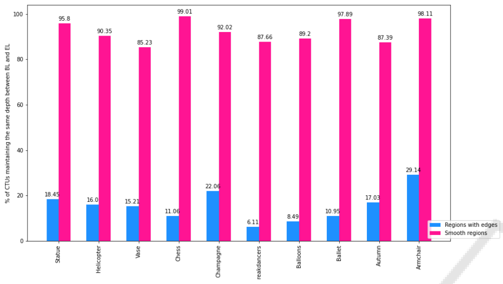


Figure 5: Percentage of CTUs maintaining the same depth between base layer and enhancement layer.

Table 1: Percentage of CTUs maintaining the same depth between base layer and enhancement layer.

Sequences	Regions with edges	Smooth regions
Statue	18.45%	95.80%
Helicopter	16.00%	90.35%
Vase	15.21%	85.23%
Chess	11.06%	99.01%
Champagne	22.06%	92.02%
Breakdancers	6.11%	87.66%
Balloons	8.49%	89.20%
Ballet	10.95%	97.89%
Autumn	17.03%	87.39%
Armchair	29.14%	98.11%

4.2 Coding Tree Unit Structure: Texture Images vs. Depth Maps

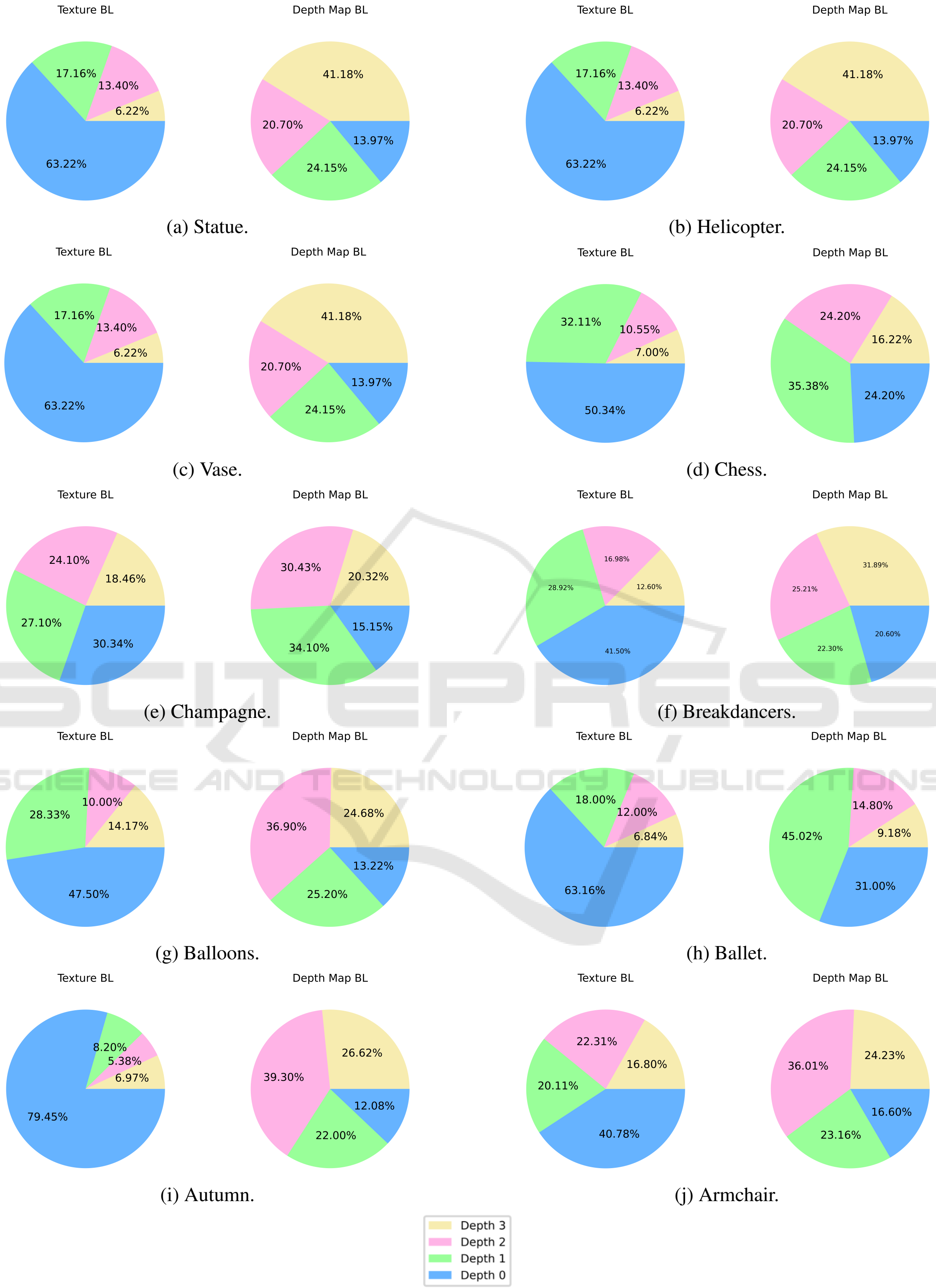
In order to study the most commonly used CTU depths (CTU Depths) in SHVC quadtree partitioning, we present the percentage of each CTU depth, i.e. 0, 1, 2 and 3, for both texture and depth images in BL (*cf.* Fig. 6) and EL (*cf.* Fig. 7). According to obtained results, we note that CTU Depths 2 and 3 are more frequently adopted, with an average percentage of 71.5%, in the texture images. They are, however, less frequently adopted in depth maps, with an average percentage of 46.74%. As texture images contain many and various details and patterns, the quadtree decomposition is needed to be partitioned to small sized coding units, e.g. 16×16 and 8×8. Thus, larger CTU depths, i.e. 2 and 3, are commonly reached. However, this is not the case of depth maps that mostly integrate smooth plans of distances to capture camera. Therefore, CTUs do not need to be very partitioned and high depths are not commonly reached. This implies that depth maps encoded using SHVC are generated with some missing information; and that is what can impact the quality of depth discontinuities and then the quality of synthesized views. Here, integrating one or more of the 3D-HEVC DMMs, besides the conventional SHVC prediction modes, can remedy this shortcoming. Furthermore, when comparing the percentage of CTUs of depth 3 in BL to those in EL, we note an increase from an average of 46.17% to 53.33% for texture images and 14.96% to 24.16% for depth maps. This can be explained by the decrease of the QP value from 26 to 22 when passing from BL to EL. In fact, more details need to be detected in EL than in BL. This involves using smaller CU of 8×8 size in order to achieve higher quality.

5 CONCLUSION

Image and video compression is a multidisciplinary and cross-cutting field that is at the crossroads of most innovative applications and domains and spans nearly every image-centric industry, where both compact and relevant information is needed. From healthcare to education, agriculture to manufacturing, and beyond, stakeholders have to rely on compression to help store and communicate their data. The study carried in this paper provides a statistical analysis of SHVC CTU partitioning for depth maps SNR scalable compression. We typically look for CTU depths in SHVC base and enhancement layers for depth maps, as well as CTU depths in texture images and their corresponding depth ones. The obtained results lead to the following conclusions:

- A high percentage of CTUs, that correspond to flat depth regions, preserves the same depth from BL to EL. However, this is not the case of non-flat depth regions.
- CTU depths reached when encoding depth maps are smaller than those reached for texture images. This can badly affect the depth discontinuities preservation, and consequently the synthesized views quality.

These interpretations can be very useful for our future research directions that aim to adapt SHVC, originally conceived for texture, to depth.



(a) Statue.

(b) Helicopter.

(c) Vase.

(d) Chess.

(e) Champagne.

(f) Breakdancers.

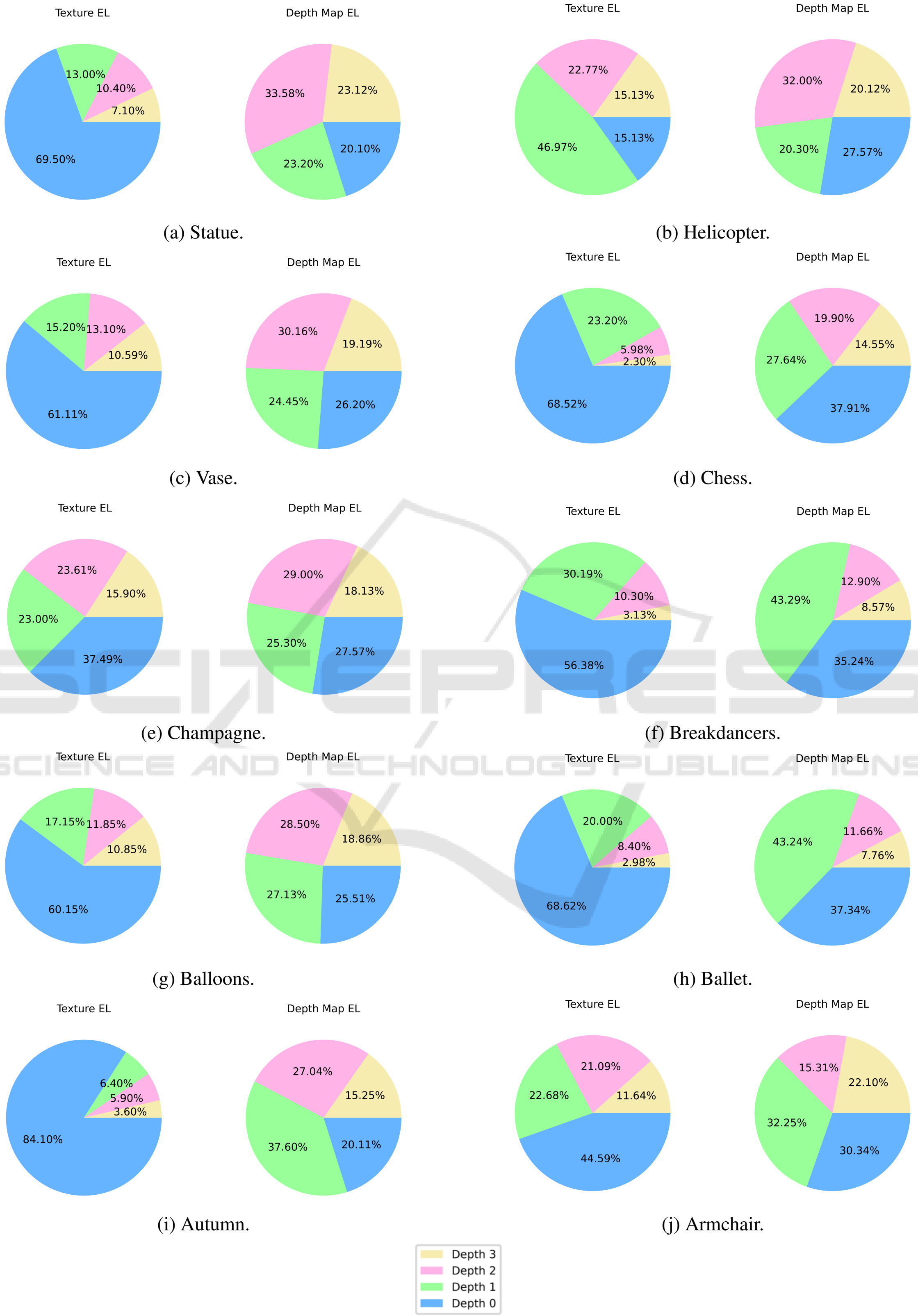
(g) Balloons.

(h) Ballet.

(i) Autumn.

(j) Armchair.

Figure 6: Percentage of CU depths for both texture images and depth maps in the base layer.



(a) Statue. (b) Helicopter.

(c) Vase. (d) Chess.

(e) Champagne. (f) Breakdancers.

(g) Balloons. (h) Ballet.

(i) Autumn. (j) Armchair.

Figure 7: Percentage of CU depths for both texture images and depth maps in the Enhancement layer.

REFERENCES

Boyce, J. M., Ye, Y., Chen, J., and Ramasubramonian, A. K. (2016). *Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard*, volume 26. IEEE Transactions on Circuits and Systems for Video Technology.

Gudumasu, S., He, Y., Ye, Y., and Xiu, X. (2015). *Video streaming with SHVC to HEVC transcoding*. SPIE, Applications of Digital Image Processing.

Merkle, P., Smolic, A., Muller, K., and Wiegand, T. (2007). *Multi-View Video Plus Depth Representation and Coding*. IEEE International Conference on Image Processing.

Sebai, D. (2020). *Performance analysis of hevc scalable extension for depth maps*, volume 92. Springer Journal of Signal Processing Systems for Signal, Image, and Video Technology.

Sullivan, G. J., Ohm, J., Han, W., and Wiegand, T. (2012). *Overview of the High Efficiency Video Coding (HEVC) Standard*, volume 22. IEEE Transactions on Circuits and Systems for Video Technology.

Sullivan, G. J. and Wiegand, T. (1998). *Rate-distortion optimization for video compression*, volume 15. IEEE Transactions on Circuits and Systems for Video Technology.

Tech, G., Chen, Y., Müller, K., Ohm, J., Vetro, A., and Wang, Y. (2016). *Overview of the Multiview and 3D Extensions of High Efficiency Video Coding*, volume 26. IEEE Transactions on Circuits and Systems for Video Technology.

Tohidypour, R. (2016). *Complexity reduction schemes for video compression*. dissertation doctorale.

Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A. (2003). *Overview of the H.264/AVC video coding standard*, volume 13. IEEE Transactions on Circuits and Systems for Video Technology.

Yea, S. and Vetro, A. (2009). *Multi-layered coding of depth for virtual view synthesis*. Picture Coding Symposium.