INTEGRAL IMAGES COMPRESSION SCHEME BASED ON VIEW EXTRACTION

A. Dricot [1,2], *J. Jung* [1]

[1] Orange Labs

M. Cagnazzo [2], *B. Pesquet* [2], *F. Dufaux* [2]

[2] Institut Mines-Télécom;
Télécom ParisTech; CNRS LTCI

ABSTRACT

Integral imaging is a glasses-free 3D video technology that captures a light-field representation of a scene. This representation eliminates many of the limitations of current stereoscopic and autostereoscopic techniques. However, integral images have a large resolution and a structure based on micro-images which is challenging to encode. In this paper a compression scheme for integral images based on view extraction is proposed. Average BD-rate gains of 15.7% and up to 31.3% are reported over HEVC. Parameters of the proposed coding scheme can take a large range of values. Results are first provided with an exhaustive search of the best configuration. Then an RD criterion is proposed to avoid exhaustive search methods, saving runtime while preserving the gains. Finally, additional runtime savings are reported by exploring how the different parameters interact.

Index Terms— Integral Imaging, Plenoptic Imaging, Holoscopy, Image and Video Coding, View Extraction

1. INTRODUCTION

3D video technologies provide immersive viewing experiences. However, current technologies on the consumer market present several limitations. The use of glasses in stereoscopy causes discomfort. Moreover the conflict between the accommodation and the convergence distances is reported to cause headaches and eyestrain. Autostereoscopic display systems use more than two views but are still limited by unrealistic perception stimuli and cannot provide smooth motion parallax (i.e. a continuous visualization when moving in front of the display), which is a key element in the perception of depth [1].

Integral imaging is a technology based on plenoptic photography [2]. This technique provides a *light-field* representation of a scene [3], which eliminates some of the current 3D technologies drawbacks (e.g. the vergence-accommodation conflict). Several companies have already been working on light-field display systems [4], which are glasses-free systems that provide a realistic visualization with smooth motion parallax. Immersive telepresence is one of the main target use cases, as well as the live 3D broadcast of sport events, e.g. 2022 World Cup in Japan, that could be projected on large light-field display systems at several public viewing facilities in major cities around the world [5].

New efficient coding technologies are required to handle the large resolution of integral images and their specific structural characteristics. We propose an efficient compression scheme that exploits view extraction techniques to create a residual integral image which is encoded. This scheme is highly parameterizable, hence we propose several iterative methods to select the most efficient configuration using a rate-distortion optimization (RDO) process.

This paper is organized as follows. In Section 2, an overview of view extraction methods is given, and state of the art methods to encode integral imaging content are presented. The proposed compression scheme is described and experimental results are shown in Section 3. Conclusions are drawn in Section 4.

2. STATE OF THE ART

2.1. View extraction

Integral imaging acquisition uses a lenticular array, composed of a large number of micro-lenses, set in front of a camera device. Integral images resulting from this capture process consist of arrays of micro-images (MIs), as illustrated in Fig. 1. Each micro-lens produces one MI, and each MI contains the light information coming from several angles of view.

Several methods to extract viewpoint images (or views) from an integral image are described in [7]. The basic method extracts one patch (a square zone of pixels) from each MI, as illustrated in Figure 2 *(left)*. This method is based on the characteristics of the focused plenoptic camera [8] for which there are both angular and spatial information within one MI. The angle of view depends on the relative position of the patch within the MI. A more basic method consists in using

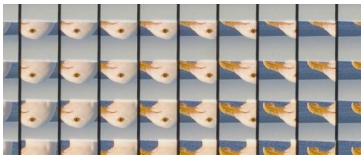


Fig. 1. Close-up on micro images (MIs) - *Seagull [6]*

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

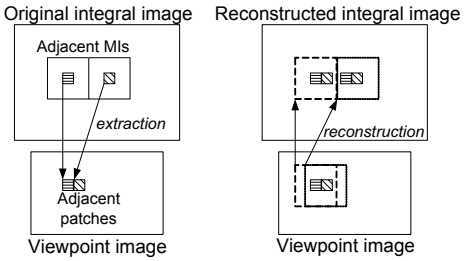


Fig. 2. View extraction (*left*), reconstruction process (*right*)

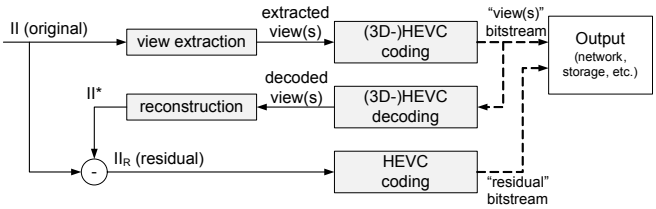


Fig. 3. Proposed scheme - *encoder side*

a patch of size 1×1, i.e. one pixel per MI. The size of the patch defines the depth plane in the scene on which the extracted view will be focused: the larger the patch, the closer the focus plane. A more advanced method allows reducing block artifacts by smoothing the transitions between adjacent patches. Pixels outside the patches' borders are blended by a weighted averaging (pixels that are further from the center have a smaller weight). A disparity estimation method (based on block matching) is proposed in [8] to obtain the depth of the objects inside each MI. It provides one estimated disparity value per MI corresponding to the adequate patch size to be used. Viewpoint images resulting from a disparity-assisted patch blending extraction (DAPBe [7]) are full-focused, as each patch size is adapted to the depth of the objects.

2.2. Integral image coding

Integral images have a large resolution in order to provide a large number of viewpoint images with a sufficient resolution. The micro-images (MIs) based structure (grid-like) involves a large number of edges which is challenging to encode (see Fig. 1). A natural approach consists in applying the Discrete Cosine Transform (DCT) to the MIs, followed by quantization and lossless coding [9]. Inter-MIs correlation can also be removed using the 3D-DCT on MIs stacked in 3D structures [10]. In [11], a Discrete Wavelet Transform (DWT) is applied to the MIs and a DCT is applied to the resulting blocks of coefficients (hybrid 4-dimensions transform). Transform-based approaches fits the MIs based structure but provide limited compression gain compared to current standard encoders (H.264/AVC [12] and HEVC [13]). In [14] and [15], integral images are decomposed in viewpoint images which are encoded using MVC encoder [16]. This approach is efficient on computer generated images (i.e. with MIs perfectly aligned on pixels) but is limited for natural integral images. Self-similarity (SS) mode [17] is another approach based on the same principle as Intra Block Copy [18], which exploits the non-local spatial correlation between MIs. A block-matching algorithm is performed as for the inter prediction modes of H.264/AVC and HEVC within the current coding frame. It provides large compression gain for still integral images but is limited for sequences when temporal prediction is enabled.

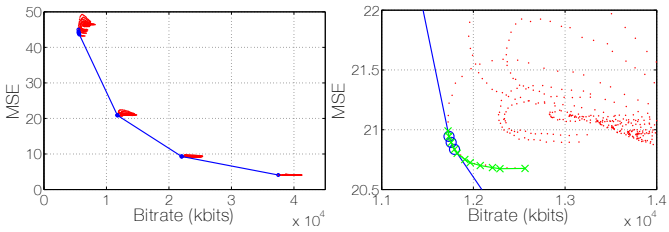
In [19] a scalable coding scheme is proposed as follows: the layer 0 corresponds to the central view, the layer 1 corresponds to a set of additional views and the layer 2 is the integral image. This layered scheme offers an interesting display scalable feature (i.e. a stream adapted to 2D, multi-view, and holoscopic display systems). However, this scalability feature comes with a cost, as additional views need to be encoded. An inter-layer prediction scheme is proposed to reduce the bitrate of layer 2, where an integral image is sparsely reconstructed from the views (layer 1) and added in the reference frame list.

In Section 3, we propose an original coding scheme to encode integral images. Although it performs view extraction, and allows some kind of display scalability, it differs from existing methods: its first goal is compression efficiency. It takes advantages of the extraction process to reconstruct a reliable predictor and create a residual integral image which is encoded.

3. PROPOSED CODING SCHEME

In this section, the proposed compression scheme (Fig. 3) is described. In this scheme, a residual integral image $\mathrm{II_R}$ is encoded with HEVC (residual stream). $\mathrm{II_R}$ is obtained by subtracting the original image II and a reconstructed image II^*. II^* is reconstructed from viewpoint images extracted from the original integral image II. Extracted views are encoded with 3D-HEVC (views stream). The number of views is not limited. Due to their small resolution, views represent a small number of bits to encode compared to II. Moreover, they have a *natural* image aspect that is less costly to encode. To obtain views with such a smooth aspect, advanced extraction methods are used (see Sec. 2.1), which use blending and patch extractions, both however preventing from perfect reconstruction. The corresponding missing information, the difference between II and II^*, is recovered in $\mathrm{II_R}$. For a reconstructed image II^* close to the original II, the subtraction provides values close to zero. Therefore $\mathrm{II_R}$ has a *flat* aspect with low variations, which is easier to encode with HEVC than II.

When reconstructing II^* from a limited number of views, some missing pixels, coming from different angles of view, are replaced by adjacent pixels from the same view, as illustrated in Fig. 2. However, the transformation of an object when changing the angle of view is not limited to a simple translation (disparity) but also involves angular differences. Hence errors are introduced. A low-pass filtering (e.g. aver-



(a) All configurations (b) Close up on $\mathrm{QP_R}$ 20

Fig. 4. Rate-distortion points for all configurations (*Fountain*)

Image	BD-rate (%)	Param. for each $\mathrm{QP_R}$ in {10,15,20,25}	
		$\mathrm{QP_V}$	B
Fountain	-17.0	19 21 23 29	3 3 3 3
Fredo	-31.3	18 21 25 32	3 3 3 3
Jeff	-5.9	25 30 30 32	9 9 9 7
Laura	-11.2	22 25 27 31	4 4 4 4
Seagull	-13.7	20 21 25 29	3 3 3 3
Sergio	-23.6	19 19 24 32	4 2 2 2
Zenhgyun1	-7.5	25 26 30 32	9 9 9 7
Average	-15.7		

Table 1. BD-Rate gains with best configurations $\mathrm{QP_V}$ and B for each $\mathrm{QP_R}$. Negative values are gains over the reference

age filter) is applied on the decoded views before the reconstruction to smooth the errors. High frequencies in the views are filtered while preserving the shape of the objects.

At the decoder side, the views are decoded and used to reconstruct II^*, and $\mathrm{II_R}$ is decoded and added to II^* to obtain the output decoded image.

There is a tradeoff between rate and quality of the views and the rate of $\mathrm{II_R}$. II^* must be as close as possible to II in order to minimize the cost of $\mathrm{II_R}$, without increasing too much the cost of the views. Several combinations are possible for the following parameters: the QP used to encode the views ($\mathrm{QP_V}$), the QP used to encode the residual image ($\mathrm{QP_R}$), and the size (in pixels) of the average filter (blur mask) applied to the decoded views (B). In the following we explore several methods to tune these parameters trading-off rate-distortion performance and complexity.

3.1. Exhaustive search of optimal configuration

Seven still images [6] listed in Table 1 are used in our experiments. Images were cropped to remove incomplete MIs and cleaned from grid pixels corresponding to the boundaries of the micro-lenses [20]. The disparity-assisted patch blending method [7] is used for the extraction of one single view. Encodings of the view and the residual image are performed under HEVC reference software (HM14.0) using the *Intra main* configuration [21], and disparity values are coded with 4 bits per MI. Compression results are provided using the Bjøntegaard Delta (BD) rate metric [22]. The anchor is II encoded with HEVC on the QP range {25,30,35,40}, and negative values represent improvement over the anchor.

As in practice most of the bitrate is dedicated to $\mathrm{II_R}$, the value of $\mathrm{QP_R}$ is set according to the target bitrate (or quality), and $\mathrm{QP_V}$ and B are considered as parameters to optimize for a given $\mathrm{QP_R}$. For each $\mathrm{QP_R}$ in the range {10,15,20,25}, combinations of values for the parameters $\mathrm{QP_V}$ and B (in the ranges {10,11,...,50} and {1,2,...,11} respectively) are iteratively tested, providing 1804 ($4 \times 41 \times 11$) rate-distortion (RD) points (red dots in Fig. 4). For each image, Table 1 shows the configuration that provides the best BD-rate results.

An average BD-rate gain of 15.7% (up to 31.3% for *Fredo*) is reported when using optimal parameter combinations. $\mathrm{QP_V}$ values increase according to $\mathrm{QP_R}$, providing a tradeoff between the bitrate for the views and for $\mathrm{II_R}$. Approximately 97% of the total bitrate is dedicated to $\mathrm{II_R}$ in average, mainly because of its very large resolution compared to the view (e.g. for *Fountain* 6512×4880 against 960×720), which represents the remaining 3% (disparity values used for extraction and reconstruction cost only 0.3%).

Optimal values for $\mathrm{QP_V}$ and B depend on the tested image and are selected among the 1804 points, which is overly complex. Results provided by this preliminary study are used in Sec. 3.2 to determine a rate-distortion optimization (RDO) process that selects the best values for a given $\mathrm{QP_R}$.

3.2. Local rate-distortion optimization process

Fig. 4 illustrates the RD values provided by 1804 different parameter combinations. We define the global convex hull (GCH, illustrated in blue) as the convex hull of all points, and the local convex hull (LCH, illustrated in green) as the convex hull of the set of 451 points with a same $\mathrm{QP_R}$ value. For a given $\mathrm{QP_R}$, optimal points are located at the intersection S of LCH and GCH. It can be observed that this intersection is not empty. GCH cannot be plotted without encoding the image with the 4 $\mathrm{QP_R}$ values, and Fig. 4 shows that using only the LCH can provide sub-optimal configurations.

The idea in this section is to be able to select the configuration (among 451 for a given $\mathrm{QP_R}$) that provides rate and distortion values (R and D respectively) minimizing a cost $D+\lambda R$, where λ is the slope of LCH in S (hence of GCH). In Fig. 4, this is equivalent to find among the points marked by a cross the points that are also marked by a circle. The slope of GCH between two consecutive values of $\mathrm{QP_R}$ (illustrated by the blue segments in Fig. 4) increases exponentially according to $\mathrm{QP_R}$. Hence an estimation of $\lambda = f(\mathrm{QP_R})$ is possible. The function defined by $\lambda = 2^{a\mathrm{QP_R}+b}$ (with $a = 0.34$ and $b = -15.8$) has an excellent fit with the data.

A first method is proposed, namely $\mathrm{All_{RDO}}$, where the 451 combinations of $\mathrm{QP_V}$ and B are processed for a given $\mathrm{QP_R}$. The combination that provides rate and distortion val-

Image	BD-rate (%)	Coding time (%)	Param. for each $\mathrm{QP_R}$ in {10,15,20,25}	
			$\mathrm{QP_V}$	B
Fountain	-17.0	48284	19 21 23 27	3 3 3 3
Fredo	-31.1	47067	18 21 25 28	3 3 3 3
Jeff	-5.9	48729	25 30 30 32	9 9 9 7
Laura	-11.2	49065	22 25 27 30	4 4 4 4
Seagull	-13.7	48836	19 21 25 29	3 3 3 3
Sergio	-23.5	48036	20 21 24 28	4 2 2 2
Zenhgyun1	-7.5	48554	25 26 31 30	9 9 9 7
Average	-15.7	48367		

Table 2. BD-Rate gain, coding time variations and associated configurations for method $\mathrm{All_{RDO}}$

ues minimizing the cost $D+\lambda R$ (with λ determined above) is selected. Test conditions are as described in Section 3.1. Table 2 shows the BD-rate and coding time variations (for each image and in average) in reference to the anchor.

Combinations selected by $\mathrm{All_{RDO}}$ are very close to the best configurations determined in Section 3.1 (same B values and only slight differences for a few $\mathrm{QP_V}$ values), and BD-rate gains of 15.7% are preserved, which shows the robustness of the estimation of $\lambda = f(\mathrm{QP_R})$. The total encoding runtime for all the iterations is large, with a multiplication of the anchor encoding time by 484 in average. It should be noted that the ranges of tested values for $\mathrm{QP_V}$ and B are not fully used and can be tightened to decrease the number of iterations. Two variant methods are proposed in the following in order to further reduce this number.

In $\mathrm{B_{RDO}}$ and $\mathrm{B_{MSE}}$, the iterations on B values are processed only for one $\mathrm{QP_V}$ value (e.g. for $\mathrm{QP_V} = 10$ in our experiment), and the best B value is kept for further $\mathrm{QP_V}$ iterations. The best value for B is the one that minimizes the cost $D + \lambda R$ in $\mathrm{B_{RDO}}$ (same RDO process as for $\mathrm{All_{RDO}}$), and the one that minimizes the mean square error (MSE) of II^* against II in $\mathrm{B_{MSE}}$. Table 3 shows the average BD-rate results and average coding time variations for these methods.

For $\mathrm{B_{RDO}}$, the total encoding runtime is significantly reduced (from 484 times the anchor down to 55 times) because the number of iterations is reduced to 51 (instead of 451 with $\mathrm{All_{RDO}}$). BD-rate gains of 15.7% are preserved because B does not vary significantly according to $\mathrm{QP_V}$. Results for $\mathrm{B_{MSE}}$ show that the encoding runtime can be further reduced (down to 44 times) by selecting B without encoding the residual image for each iteration, with an average BD-rate gain almost as good (15.3% in average, e.g. with a decrease of 1.7% for *Seagull*, and 0.8% for *Sergio*). It should be noted that the number of iterations on $\mathrm{QP_V}$ can be further reduced by avoiding the full search on the range {10,11,...,50}. For example, it can generally be observed that the cost $D + \lambda R$ has one local minimum according to $\mathrm{QP_V}$, for B and $\mathrm{QP_R}$ given. Hence the iterations on $\mathrm{QP_V}$ can stop when the cost starts to increase.

Method	BD-Rate (%)	Coding time (%)
$\mathrm{All_{RDO}}$	-15.7	48367
$\mathrm{B_{RDO}}$	-15.7	5526
$\mathrm{B_{MSE}}$	-15.3	4443
$\mathrm{QPV_{fixed}}$	-15.5	136
$\mathrm{All_{fixed}}$	-8.5	120

Table 3. Average BD-Rate gains and coding time variations

3.3. Empirical selection of parameter values

We define two variants, $\mathrm{QPV_{fixed}}$ and $\mathrm{All_{fixed}}$, where a value of $\mathrm{QP_V}$ is empirically fixed for each $\mathrm{QP_R}$. In $\mathrm{QPV_{fixed}}$, B is iteratively selected according to the MSE of II^* against II (as described in Sec. 3.2), while in $\mathrm{All_{fixed}}$, B is also fixed.

In Tab. 3, results for $\mathrm{QPV_{fixed}}$ show that assigning one $\mathrm{QP_V}$ to one $\mathrm{QP_R}$ largely reduces the encoding runtime (only 1.4 times the anchor) and still provides 15.5% BD-rate gains in average, which is close to optimal. Although the number of available images is limited, parameters only slightly differ from one image to another. This robustness suggests similar gains on other integral images. For $\mathrm{All_{fixed}}$, the coding time is only 1.2 times the anchor time. However, the BD-rate gain drops to 8.5%, with losses for *Jeff* and *Zenhgyun1*. The adequate B value strongly depends on the image and iterations on this parameter can significantly improve the coding performance, with only a slight increase of the encoding runtime.

Preliminary results show similar performances between the proposed scheme and the Intra Block Copy [18] mode (same principle as the state-of-the-art SS [17] method), and that combining both methods provides increased efficiency.

Table 4 shows the encoding and decoding runtime variations against the anchor for *Fountain* with the $\mathrm{QPV_{fixed}}$ method, and the percentage of the runtime dedicated to each task. Encoding runtime is 1.3 times the encoding time of II with HEVC (anchor), with encoding of $\mathrm{II_R}$ representing 79% of the total time. The eleven iterations of blur, reconstruction, and subtraction steps represent 12%. View extraction represents 7%, mainly because of the time-consuming disparity estimation. Decoding runtime does not depend on the number of iterations at the encoder side. It is 2.4 times the anchor, with 46% for the decoding of $\mathrm{II_R}$. Reconstruction (31%) and sum (22%) represent a larger percentage at the decoder

Runtime (%)	against anchor	Extr.	Rec.	HEVC		Others
				View	$\mathrm{II_R}$	
Encoding	130	7	8	2	79	4
Decoding	240	/	31	1	46	22

Table 4. *Fountain* - Runtime variation against anchor with $\mathrm{QPV_{fixed}}$, and percentage of the total time for each task including: extraction, reconstruction, view and residual encoding/decoding, and blur, subtraction and sum as *others*.

because HEVC decoding process is much faster than encoding. The increase is larger in lower bitrates where HEVC decoding time is further reduced while the reconstruction and sum runtime do not vary.

4. CONCLUSIONS

In this paper we propose an efficient integral image compression scheme where a residual integral image and an extracted view are encoded. The residual image is the difference between the original image and an image reconstructed from the view. An average BD-rate gain of 15.7% up to 31.3% over the HEVC anchor is reported. Coding performance largely depends on the configuration of the QP used to encode the view and the size of a low-pass filter applied to the view. A robust iterative RDO process is modeled to select the best configuration, preserving optimal BD-rate gains. We show that the number of iterations can be limited to reduce the runtime while preserving BD-rate gains. Finally we prove that we can assign one single QP for the view to a given QP for the residual with minimal loss, and that the low-pass filter size can be selected using reduced iterations. This results in a realistic coding performance vs runtime codec. In future work, the coding scheme will be tested with several views extracted, with different filtering for the view(s), and with more advanced extraction methods.

REFERENCES

[1] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging technologies for 3D video: content creation, coding, transmission and rendering*, Wiley Eds, 2013.

[2] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[3] A. Lumsdaine and T. Georgiev, "Full resolution lightfield rendering," *Indiana University and Adobe Systems, Tech. Rep*, 2008.

[4] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, and T. Fujii, "[m31261][FTV AHG] Multiple aspects," in *ISO/IEC JTC1/SC29/WG11*, October 2013.

[5] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, and T. Fujii, "[m31103][FTV AHG] Introduction of super multiview video systems for requirement discussion," in *ISO/IEC JTC1/SC29/WG11*, October 2013.

[6] "http://www.tgeorgiev.net/," .

[7] J. F. O. Lino, "2D image rendering for 3D holoscopic content using disparity-assisted patch blending," *Thesis to obtain the Master of Science Degree*, October 2013.

[8] T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021106, 2010.

[9] M. Forman, A. Aggoun, and M. McCormick, "Compression of integral 3D TV pictures," in *Fifth International Conference on Image Processing and its Applications*, Edinburgh, UK, July 1995, IET, pp. 584–588.

[10] M. C. Forman, A. Aggoun, and M. McCormick, "A novel coding scheme for full parallax 3D-TV pictures," in *ICASSP Proceedings*, Nagoya, Japan, August 1997, IEEE, vol. 4, pp. 2945–2947.

[11] E. Elharar, A. Stern, O. Hadar, and B. Javidi, "A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform," *Journal of display technology*, vol. 3, no. 3, pp. 321–325, September 2007.

[12] D. Marpe, T. Wiegand, and G. J. Sullivan, "The H. 264/MPEG4 advanced video coding standard and its applications," *Communications Magazine*, vol. 44, no. 8, pp. 134–143, 2006.

[13] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *TCSVT*, vol. 22, no. 12, pp. 1649–1668, 2012.

[14] J. Dick, H. Almeida, L. D. Soares, and P. Nunes, "3D holoscopic video coding using MVC," in *EUROCON*, Lisbon, Portugal, April 2011, IEEE, pp. 1–4.

[15] S. Shi, P. Gioia, and G. Madec, "Efficient compression method for integral images using multi-view video coding," in *18th ICIP*, Brussels, Belgium, September 2011, IEEE, pp. 137–140.

[16] J.-R. Ohm, "Overview of 3D video coding standardization," in *3DSA*, Osaka, Japan, June 2013.

[17] C. Conti, P. Nunes, and L. D. Soares, "New HEVC prediction modes for 3D holoscopic video coding," in *19th ICIP*, Orlando, FL, September 2012, IEEE, pp. 1325–1328.

[18] D.-K. Kwon and M. Budagavi, "Fast intra block copy (intrabc) search for hevc screen content coding," in *ISCAS*. IEEE, 2014, pp. 9–12.

[19] C. Conti, P. Nunes, and L. D. Soares, "Inter-layer prediction scheme for scalable 3-d holoscopic video coding," *Signal Processing Letters*, vol. 20, no. 8, pp. 819–822, 2013.

[20] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoffs in integral photography.," *Rendering Techniques*, vol. 2006, pp. 263–272, 2006.

[21] F. Bossen, "Test conditions and software reference configurations," *JCT-VC L1100*, January 2013.

[22] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," in *VCEG Meeting*, Austin, USA, April 2001.