# A STUDY ON THE 4D SPARSITY OF JPEG PLENO LIGHT FIELDS USING THE DISCRETE COSINE TRANSFORM

*Gustavo Alves*, Márcio P. Pereira*, Murilo B. de Carvalho§*, Fernando Pereira‡,*
*Carla L. Pagliari*†, Vanessa Testoni** and Eduardo A. B. da Silva**

*PEE/COPPE/DEL/POLI/UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
†PEE/PGED/IME, Instituto Militar de Engenharia, Rio de Janeiro, Brazil
‡Instituto Superior Técnico, Universidade de Lisboa - Instituto de Telecomunicações, Lisbon, Portugal
§TET/CTC, Universidade Federal Fluminense, Niterói, Brazil
**Samsung Research Brazil

## ABSTRACT

In this work we study the 4D sparsity of light fields using as main tool the 4D-Discrete Cosine Transform. We analyze the two JPEG Pleno light field datasets, namely the lenslet-based and the High-Density Camera Array (HDCA) datasets. The results suggest that the lenslets datasets exhibit a high 4D redundancy, with a larger inter-view sparsity than the intra-view one. For the HDCA datasets, there is also 4D redundancy worthy to be exploited, yet in a smaller degree. Unlike the lenslets case, the intra-view redundancy is much larger than the inter-view one. The results and conclusions of this first study for light field imaging may have a strong impact on the current and future design of efficient codecs for this type of emerging data, notably in the context of the JPEG Pleno standard.

***Index Terms***— Light field, coding, sparsity, 4D-DCT

## 1. INTRODUCTION

Image-based representation models are an alternative to geometry-based representation models where light modeling uses the plenoptic illumination function, which represents everything visible from any point in the 3D space [1]. A light field (LF) is a simplification of the plenoptic function when the intensity along a light ray is constant. Thus, in a LF a light ray can be parameterized by its intersection with two planes, yielding a 4D function if time is not considered [2, 3]. It can be acquired by an array of cameras or by a single camera with special lenses (lenslet-based), naturally offering different fields of view. In contrast to stereo systems, the 4D structure of the LF enables the accurate reconstruction of multiple viewpoints in the 3D world, as well as other operations requiring manipulation of the light rays, such as refocusing. In order for such properties of LFs to be fully explored, the density of light rays should be large, which is equivalent to having a large number of viewpoints. This usually generates a massive amount of data. Both in the lenslet-based and the camera array-acquired cases, the LFs can be represented as 2D arrays of regular images corresponding to viewpoints distributed on a 2D grid. [1-3]. Clearly, an efficient coding scheme is essential to reduce this large amount of data for storage and transmission. To this end, one wants to effectively explore the redundancy present in LFs. Besides the well-known redundancy contained within each view or subaperture image of a LF, a large amount of redundancy can also be observed when analyzing LF epipolar plane images (EPIs), that are images assembled using one spatial and one view dimension. There, one can see that patches in any of the views are slightly shifted in all neighboring views, implying that there is also a large amount of space-view redundancy [4].

In [5], this 4D redundancy is investigated by using a 4D discrete cosine transform (DCT) basis in a compressive sensing framework. The obtained results suggest that the 4D-DCT has the potential of being an effective tool for LFs compression. Let's consider that a light field captured scene is made up of Lambertian objects at known depths. Then, it is reasonable to assume that, given a single image (view) and its corresponding depth map, it is possible to reconstruct all 4 dimensions of the LF [6]. In this paper the term sparsity is used in the sense of [6], that is related to how much of the energy of the signal is concentrated, for a given s, in the s% transform coefficients with largest variances.

Such need for efficient LF coding schemes is driving standardization activities, notably from JPEG, that has issued the JPEG Pleno Call for Proposals on Light Field Coding in January 2017 [7-9]. It requests contributions for coding solutions in the area of LFs encompassing both lenslet-generated LF images and LF images obtained using high density 2D camera arrays (HDCA).

From the above, considering the four-dimensional (4D) nature of the LF data, novel coding solutions that are able to explore the available 4D redundancy should be developed. One natural approach to this end is the extension of the 2D coding tools used for regular 2D images to 4D LF data. Thus, 4D transforms are natural candidates for tools that can properly explore the full 4D redundancy. In this context, this paper proposes to use a 4D transform (the 4D-DCT), in order to investigate the 4D sparsity of the LFs. We believe that such a study can potentially impact the current and future design of LF coding solutions, notably within JPEG Pleno. This is particulary relevant considering that most of the available LF coding solutions, including those currently under development in JPEG Pleno, do not fully take into account the intrinsic 4D nature of such imaging data.

The remainder of this paper is organized as follows. Section 2 describes the JPEG Pleno datasets used for the proposed study, while the experimental framework is detailed in Section 3. The results are discussed in Section 4 and Section 5 presents the final remarks.

## 2. JPEG PLENO DATASETS

The LFs used in this study are described in the JPEG Pleno Call for Proposals on Light Field Coding [7]. The JPEG Pleno LFs include two datasets acquired by different devices and setups: lenslet-based plenoptic cameras and high-density array of conventional cameras (HDCA).

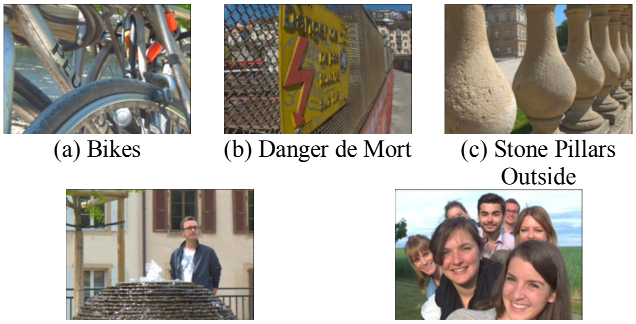
978-1-4799-7061-2/18/$31.00 ©2018 IEEE

### 2.1. Lenslet-based light field dataset

Fig. 1 shows the central views of the LFs acquired with a lenslet based camera, Lytro Illum B01 (10-bit) [10]. Each LF consists of a 15×15 array of views, with spatial resolution of 626×434 pixels each. The images selected from this dataset, pictured in Fig. 1, are natural outdoor images presenting different levels of spatial information, with objects at different depths and repetitive patterns [11].

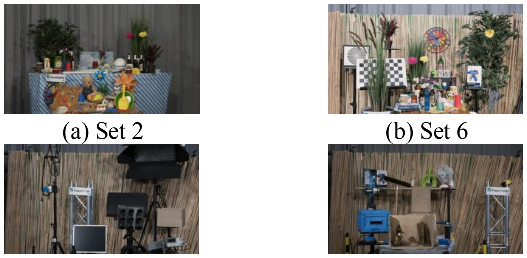
### 2.2. HDCA light field dataset

The central views of the HDCA LFs are depicted in Fig. 2. They were acquired using a high resolution camera attached to a moving robot. All images portray rather similar indoor (studio) scenes and show different levels of detail, specularities, regular patterns and objects at different depths. Each LF has 101×21 views each with a 3840×2160 spatial resolution. This study also uses the view-subsampled version of the HDCA datasets, notably 33 × 11 views, with the same spatial resolution, as defined in the JPEG Pleno core experiments [12]. While this view-sampled of the HDCA context has less inter-view redundancy, its usage within JPEG Pleno resulted from the need to limit the computational complexity associated to the HDCA content coding.


(a) Bikes (b) Danger de Mort (c) Stone Pillars Outside

(d) Fountain & Vincent 2 (e) Friends 1

**Fig. 1:** Selected lenslet light field images (central views).


(a) Set 2 (b) Set 6

(c) Set 9 (d) Set 10

**Fig. 2:** Selected HDCA light field images (central views).

## 3. EXPERIMENTAL FRAMEWORK

This LFs sparsity study is based on a simple experimental framework with the processing pipeline depicted in Fig. 3. As a first step, a 4D data block is extracted from the input LF organized as a 2D array of views (for the lenslet LF images, these views are the subaperture images). In this work we use the variables $s$ and $t$ for the view coordinates and $u$ and $v$ for the image coordinates within each $(s, t)$ view. Next, a separable 4D-DCT (Fig. 4) is applied to each 4D LF block followed by a thresholding operation on the 4D-DCT coefficients, which basically performs a selection of the coefficients based on their energy. The 4D-DCT coefficients with values higher than a pre-defined threshold are retained, while the others are discarded. Naturally, the lower this threshold, the more coefficients are selected. The retained coefficients are used to reconstruct the 4D data blocks by using the inverse 4D-DCT. Finally, the full LF image is (lossy) recovered by appropriately combining the reconstructed blocks. The more coefficients are retained, the higher the quality of the recovered LF.

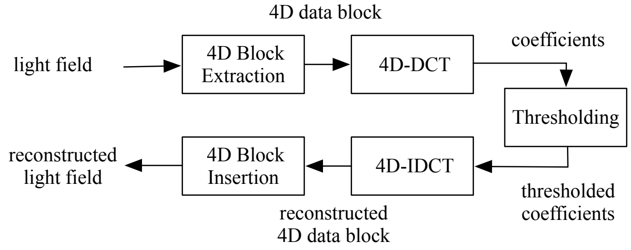


**Fig. 3:** Experimental framework processing pipeline.

Fig. 4 illustrates de the 4D-DCT processing pipeline. In this study, 4D-DCT sizes ($t \times s \times v \times u$) used for the lenslet LF images are 8×8×8×8 (4D), 1×1×8×8 (2D intra-view) and 8×8×1×1 (2D inter-view). While the 8×8×8×8 4D block size allows exploiting the 4D redundancy, the same does not happen for the 1×1×8×8 and 8×8×1×1 block sizes which exploit only the intra or only the inter-view redundancies, respectively. In order not to mix different block sizes in one experiment. both inter-view and intra-view dimensions were truncated to the nearest multiple of 8. We have chosen to use only the central 624×432 portion of each view and only the central 8×8 subaperture views. This choice of subaperture views had the desirable consequence of avoiding to use the darkened views associated to the vignetting effect.

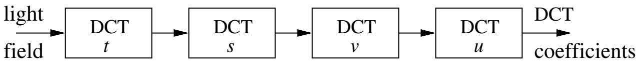


**Fig. 4:** 4D-DCT processing pipeline.

For the HDCA LF datasets, the same 4D-DCT sizes are used, with the addition of the 8×8×64×64 4D-DCT (8×8 inter-view and 64×64 intra-view). The arrays of views are also truncated to the nearest multiple of 8, thus resulting in the choice of the central 96×16 views for the original HDCA dataset and 32×8 views for the subsampled HDCA dataset version.

To further validate the experimental results and conclusions, we include in our analysis the computation of the GSVR (Geometric Space View Redundancy) descriptor [13] for the LFs analyzed. The GSVR expresses the permanence probability of the image of a point in 3D space across the views from a 4D space-view block**.** It characterizes LFs in terms of space-view redundancy. Although it does not take into account the intra-view redundancy, it may bring interesting insights into our results.

## 4. RESULTS AND ANALYSIS

This section will present the results and conclusions of the study performed both for the JPEG Pleno lenslet and HDCA LFs. The sparsity of the LFs will be assessed through the PSNR (which is equivalent to the concentration of energy) for a given concentration of energy in a specific percentage of retained 4D-DCT coefficients per 4D block (which is equivalent to the sparsity).

### 4.1. Lenslet light field datasets

Fig. 5 shows the sparsity results for the average PSNR-Y versus the average % of retained coefficients per block for the 8×8×8×8 4D-DCT and all the lenslets datasets. The charts show that the Friends dataset has more 4D sparsity than the other ones. Fig. 6 shows the same type of results for the LF Bikes when using different block sizes, notably able to exploit different types of redundancy. The sparsity is larger for the 4D-DCT than for its 2D counterparts, thus implying that it is worthy to exploit the 4D redundancy in a DCT-based coding scheme for lenslets. In addition, the inter-view sparsity is larger than the intra-view one, indicating that for the lenslets it is more effective to use an inter-view transform than an intra-view one. We have observed the same behavior for the other LFs of the lenslet dataset.

Fig. 7 displays the GSVR curves [13] for the lenslets datasets. For a permanence probability of 0.9, the GSVR curves show that the intra-view block-size can vary from 0.3 to 0.5 times the inter-view one. This confirms that the inter-view redundancy in indeed larger than the intra-view one for the lenslet LFs since one can have a larger interview block dimension for the same permanence probability.

### 4.2. HDCA light field datasets

Fig. 8 shows the sparsity results expressed by the average PSNR-Y versus the average % of retained coefficients per block for the 8×8×8×8 4D-DCT and all original (101×21) HDCA datasets. The results show that the Set 2, Set 9 and Set 10 datasets are approximately equivalent regarding the 4D-DCT sparsity, while the sparsity of Set 6 is smaller. Fig. 9 compares the average PSNR-Y versus the average % of retained coefficients per block results using different block sizes for the HDCA LF Set 2. Unlike the lenslets datasets, where the 4D sparsity is significantly larger than the 2D sparsities, for the HDCA datasets the sparsity is dominated by the intra-view redundancy. In addition, the inter-view sparsity is much smaller than the intra-view one. The same behavior is observed for the other HDCA datasets.

Fig. 10 shows the average PSNR-Y versus the average % of retained coefficients per block results for the 8×8×8×8 4D-DCT for all subsampled HDCA datasets, while Fig. 11 compares the 4D-DCT sparsity to the 2D inter-view (8×8×1×1) and 2D intra-view (1×1×8×8) sparsities for the subsampled HDCA Set 2 dataset. The behavior is very similar to the one for the corresponding original HDCA (Fig. 8 and 9) thus implying that for the subsampled HDCA datasets the sparsity is also dominated by the intra-view redundancy, with a much smaller inter-view redundancy.

Fig. 12, 13 and 14 compare the 4D-DCT, 2D intra-view DCT and 2D inter-view DCT sparsity results for the original and subsampled HDCA Set 2. The sparsity of the original dataset is much larger than for the same subsampled dataset for both the 4D and 2D inter-view DCT, with no difference for the 2D intra-view DCT. This is expected, due to the larger spacing between adjacent views, introduced by the view subsampling, with leads to a large reduction in inter-view redundancy for the subsampled HDCA dataset.

Finally, Fig. 15 compares the sparsity of the 8×8×8×8 and the 8×8×64×64 4D-DCT for both the original and subsampled HDCA datasets. The difference between these two DCT sizes is only on the intra-view dimensions (8×8 and 64×64). It is possible to observe that, for the same sparsity level, the difference in PSNR values between the two 4D-DCT sizes is higher for the original datasets. Since the difference between the two datasets is only in the inter-view redundancy, this allows concluding that the intra-view block size impacts on the exploitation of the inter-view redundancy, with larger intra-view dimensions being better. This is in accordance with the results displayed in Fig. 16, showing the GSVR [13] curves for the original and subsampled HDCA datasets. For a permanence probability of 0.8, an intra-view dimension around 8 times larger than the inter-view's one should be used for the original HDCA datasets, which, for 8×8 inter-view dimensions requires intra-view dimensions of around 64×64. On the other hand, an intra-view dimension around 20 times larger than the inter-view's dimension should be used for the subsampled HDCA datasets, which would require intra-view dimensions of around 160×160 for the 8×8 inter-view dimensions. Then, an 8×8×64×64 block would be adequate for exploring the 4D redundancy in the original HDCA datasets, while it would not be sufficient for the subsampled HDCA, that would require an 8×8×160×160 4D-DCT size. This explains the fact that the sparsity increase from 8×8×8×8 to 8×8×64×64 is much larger for the original than for the subsampled HDCA dataset. In addition, one should note that, if the intra-view dimensions are too large, the intra-view redundancy across the block becomes smaller and cannot be well exploited.

### 4.3. Lenslet versus HDCA light field datasets

From the above results on can conclude that lenslet dataset has a great deal of 4D sparsity, the inter-view redundancy being significantly larger than the intra-view. Unlike the lenslet dataset, the intra-view sparsity of the HDCA dataset is much larger than the inter-view. In addition, the HDCA datasets have a much smaller amount of 4D redundancy than the lenslet dataset. Thus, it is likely that the coding solutions that are more efficient to the lenslet datasets will not be the more efficient ones to the HDCA datasets.

## 5. FINAL REMARKS

A light field is a 4D data structure with a large amount of redundancy. The presented study of the 4D sparsity of the JPEG Pleno LF images show that both the lenslet and HDCA datasets have a great amount of 4D redundancy that can be explored, notably for coding purposes. As a consequence, the results also suggest that not exploiting the 4D redundancy as a whole may be a severe limitation to the design or emerging LF codecs, notably the one currently under development in JPEG Pleno. The presented results also suggest that the HDCA and lenslet datasets may require distinct coding solutions due to the different nature of their 4D redundancy. One should bear in mind that these conclusions are restricted to the JPEG Pleno datasets, whereas a more extensive study needs to be carried out to characterize the inter-view and intra-view redundancies of more general LF data.

## 6. ACKNOWLEDGMENT

The authors would like to thank the support of CAPES, FAPERJ, CNPq, TV Globo and Samsung Brazil Research Center.

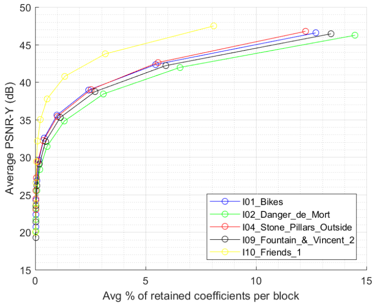


**Fig. 5:** Average PSNR-Y vs average % of retained coefficients per block for the lenslets datasets using 8×8×8×8 4D-DCT.

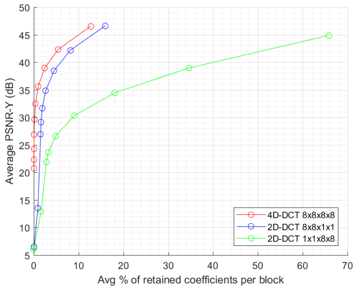


**Fig. 6:** Average PSNR-Y vs average % of retained coefficients per block for Bikes: Comparison of 8×8×8×8 4D-DCT, inter-view 8×8 2D-DCT and intra-view 8×8 2D-DCT.

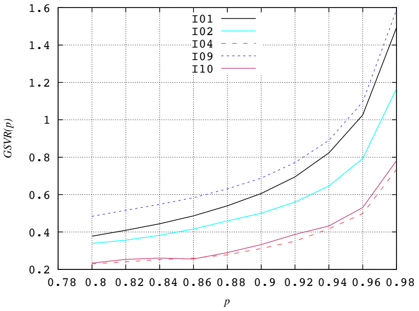


**Fig. 7:** GSVR descriptor for the lenslets datasets [13].

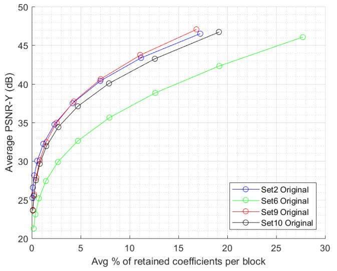


**Fig. 8:** Average PSNR-Y vs average % of retained coefficients per block for the HDCA datasets using 8×8×8×8 4D-DCT.

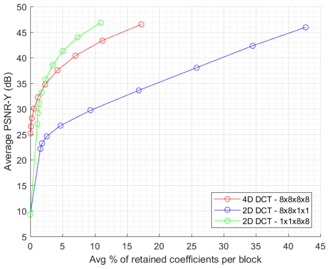


**Fig. 9:** Average PSNR-Y vs average % of retained coefficients per block for HDCA Set 2: Comp. of 8×8×8×8 4D-DCT, inter-view 8×8 2D-DCT and intra-view 8×8 2D-DCT.

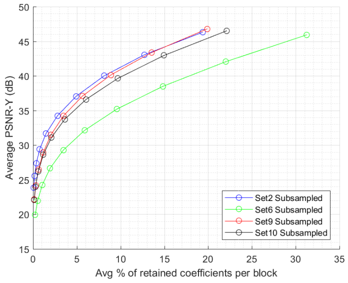


**Fig. 10:** Average PSNR-Y vs average % of retained coefficients per block for the subsampled HDCA datasets: 8×8×8×8 4D-DCT.

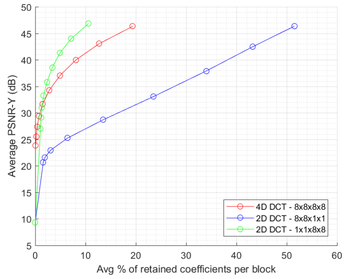


**Fig. 11:** Average PSNR-Y vs average % of retained coefficients per block for the subsampled HDCA Set 2: Comparison of 8×8×8×8 4D-DCT, inter-view 8×8 2D-DCT and intra-view 8×8 2D-DCT.

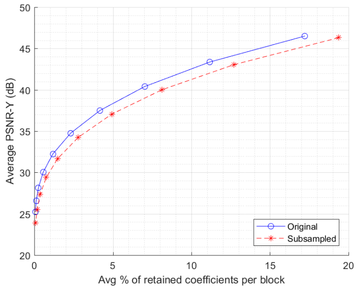


**Fig. 12:** Average PSNR-Y vs average % of retained coefficients per block for HDCA Set 2, original and subsampled: 8×8×8×8 4D-DCT.

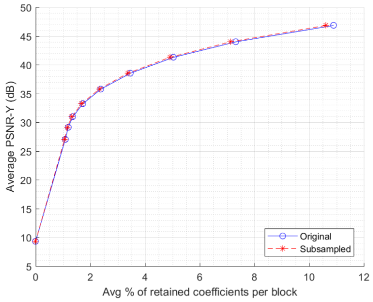


**Fig. 13:** Average PSNR-Y vs average % of retained coefficients per block for HDCA Set 2, original and subsampled: intra-view 8×8 2D-DCT.

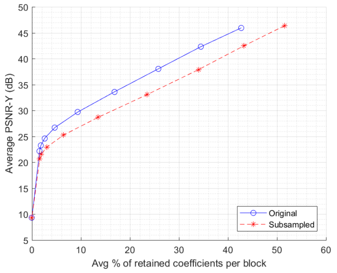


**Fig. 14:** Average PSNR-Y vs average % of retained coefficients per block for HDCA Set 2, original and subsampled: inter-view 8×8 2D-DCT.

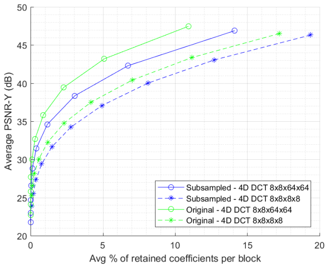


**Fig. 15:** Average PSNR-Y vs average % of retained coefficients per block for HDCA Set 2, original and subsampled: Comparison of 8×8×8×8 and 8×8×64×64 4D-DCT.

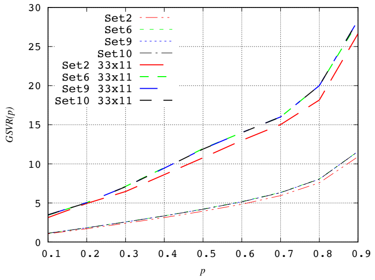


**Fig. 16:** GSVR descriptor for HDCA and subsampled HDCA datasets [13].

## 7. REFERENCES

[1] Adelson, E. H., Bergen J. R., “The Plenoptic Function and the Elements of Early Vision,” M. Landy and J. A. Movshon, (eds) Computational Models of Visual Processing, 1991.

[2] Ng, R., Digital Light Field Photography, Ph.D. Thesis, Stanford, CA, USA, 2006.

[3] Dansereau, D.G., Plenoptic Signal Processing for Robust Vision in Field Robotics, Ph.D. Thesis, University of Sydney, Australia, 2014.

[4] Johannsen O., Sulc A., and Goldluecke B., "What Sparse Light Field Coding Reveals about Scene Structure," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

[5] Miyagi Y., Takahashi K., *et al,* "Reconstruction of Compressively Sampled Light Fields Using a Weighted 4D-DCT Basis," IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, September 2015.

[6**]** Shi, L. *et al*, “Light Field Reconstruction Using Sparsity in the Continuous Fourier Domain”. ACM Transactions on Graphics. 34. 1-13. 2014.

[7] ISO/IEC JTC 1/SC29/WG1 N74014, “JPEG Pleno Call for Proposals on Light Field Coding,” Geneva, Switzerland, January 2017.

[8] “Overview of JPEG Pleno, ”https://jpeg.org/ jpegpleno/index.html.

[9] ISO/IEC JTC 1/SC29/WG1N75024, ”JPEG Pleno Call for Proposals – Submission Process Details,” Sydney, Australia, March 2017.

[10] “Lytro,” https://www.lytro.com/.

[11] Rerabek, M. and Ebrahimi, T. “New light field image dataset,” in 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016.

[12] ISO/IEC JTC 1/SC 29/WG1N77016, Core Experiments Set #2 for JPEG Pleno . Date: 2017 October 26.

[13] Pereira, M. P., Alves, G. O., Pagliari, C. L., Carvalho, M. B., da Silva, E. A. B., Pereira, F., "Geometric Space-View Redundancy Descriptor for Light Fields: Predicting the Compression Potential of the JPEG Pleno Light Field Datasets", IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, October 2017.