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

Gustavo Alves*, Marcio P. Pereira*, Murilo B. de Carvalho§, Fernando Pereira‡, Carla L. Pagliari†, Vanessa Testoni**, 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 Tecnico, Universidade de Lisboa - Instituto de Telecomunicacoes, Lisbon, Portugal - §TET/CTC, Universidade Federal Fluminense, Niteroi, Brazil - **Samsung Research Brazil

1 - Context

- Light fields (LF): 4D structures that contain the images of a given scene from a sampled 2D dense range of viewpoints;
- An efficient coding scheme for LF is essential to reduce the large amount of data for LF storage and transmission;
- The need for efficient LF coding schemes is driving standardisation activities, notably from JPEG Pleno;
- 4D transforms are natural candidates for tools that can properly explore the full LF redundancy;

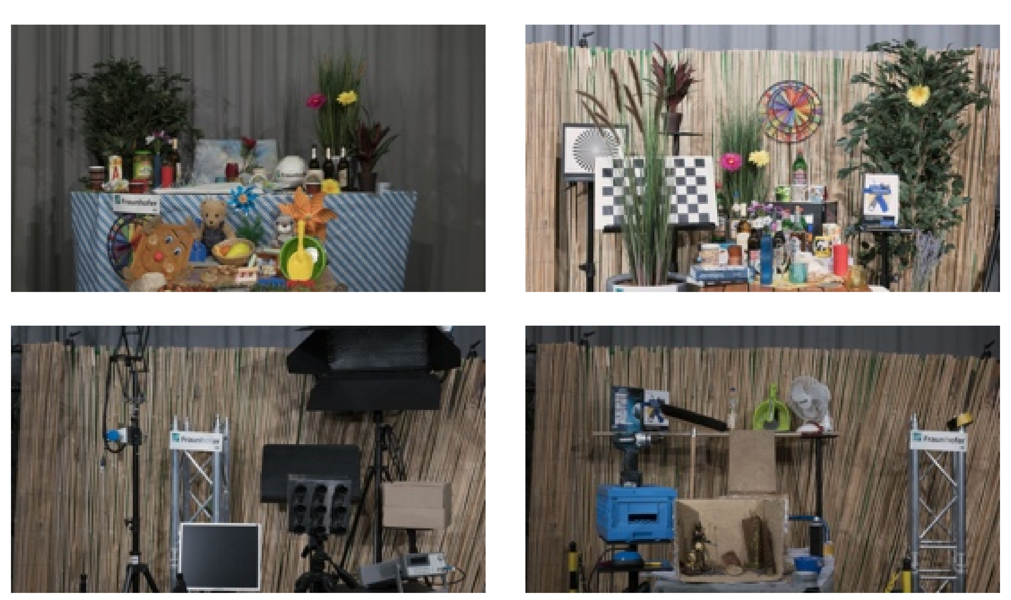
2 - Objective

- **This work proposes to use the 4D-DCT in order to investigate the 4D sparsity of the light fields;**
  - Sparsity: how much of the energy of the signal is concentrated in the $s\%$ transform coefficients with largest variances;
- Such a study can potentially impact the current and future design of LF coding solutions, notably within JPEG Pleno;

3 - JPEG Pleno Light Fields Datasets

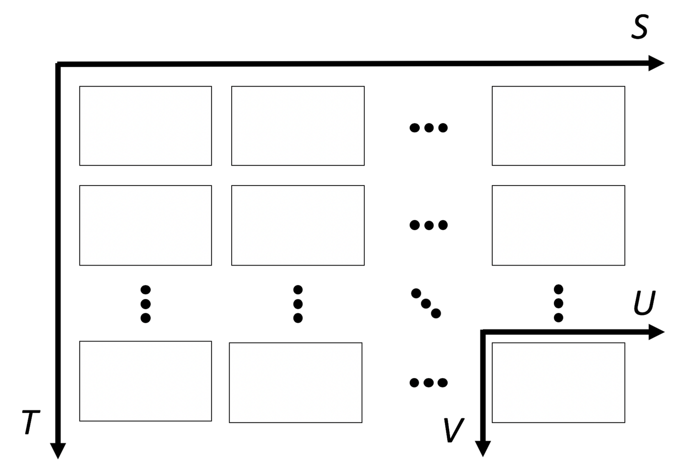
- Lenslet-based Datasets.
  - Each light field: 15×15 views with 626×434 pixels each.

- HDCA Datasets.
  - Full datasets: 101×21 views with 3840×2160 pixels each;
  - Subsampled datasets: 33×11 views with 3840×2160 pixels each.

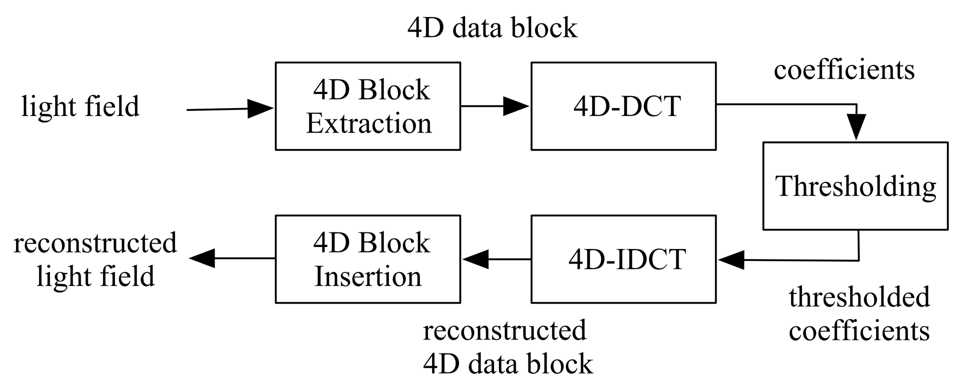


4 - Experimental Framework

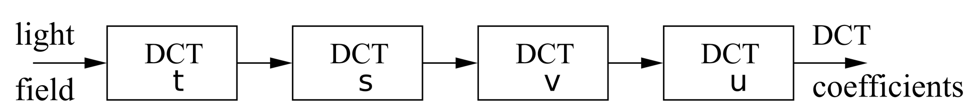
- 4D or 2D DCT is applied to (t, s, v, u) blocks:



- Experimental processing pipeline:



- Separable 4D-DCT pipeline:



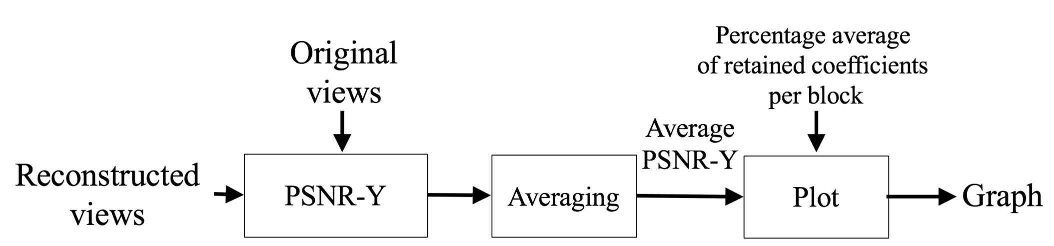
- Sizes:

	Lenslet-Based	HDCA
DCT Sizes	$8 \times 8 \times 8 \times 8$ (4D) $1 \times 1 \times 8 \times 8$ (2D intra view) $8 \times 8 \times 1 \times 1$ (2D inter view)	$8 \times 8 \times 8 \times 8$ (4D) $8 \times 8 \times 64 \times 64$ (4D) $1 \times 1 \times 8 \times 8$ (2D intra view) $8 \times 8 \times 1 \times 1$ (2D inter view)
Views Used	8×8 central views	Original: 16×96 Subsampled 8×32
Views Dimensions	432×624	2160×3840

- Only luminance is used.

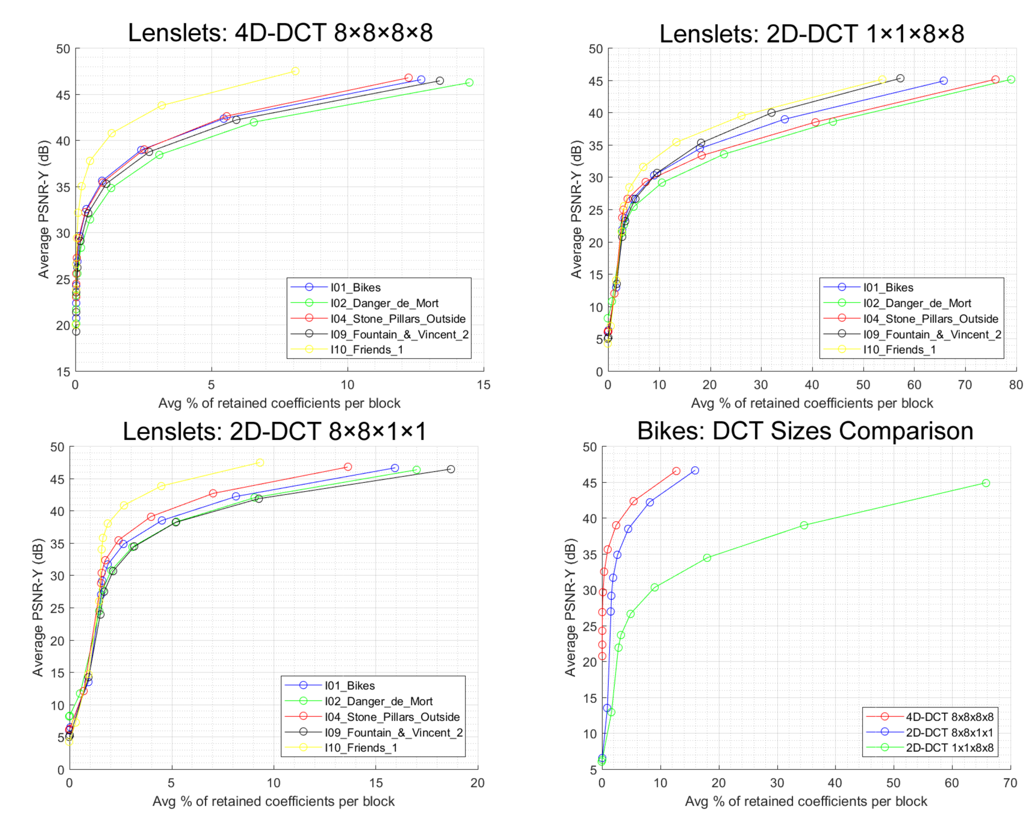
5 - Analysis

- Plot: average PSNR of the reconstructed views VS average percentage of retained coefficients per block:
  - PSNR: equivalent to the concentration of energy;
  - Percentage of retained 4D-DCT coefficients per 4D block: equivalent to the sparsity.



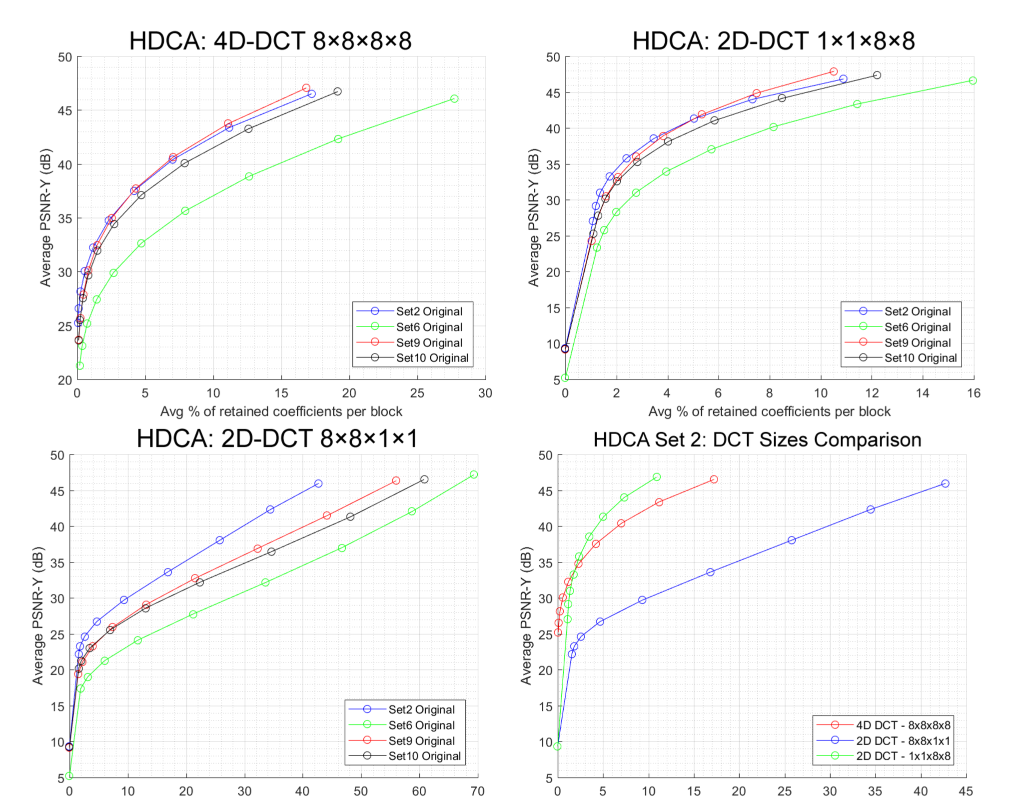
- Geometric Space View Redundancy (GSVR) descriptor:
  - Expresses the permanence probability of the image of a point in 3D space across the views from a 4D space-view block;
  - Characterises LFs in terms of space-view redundancy.

6 - Results: Lenslets



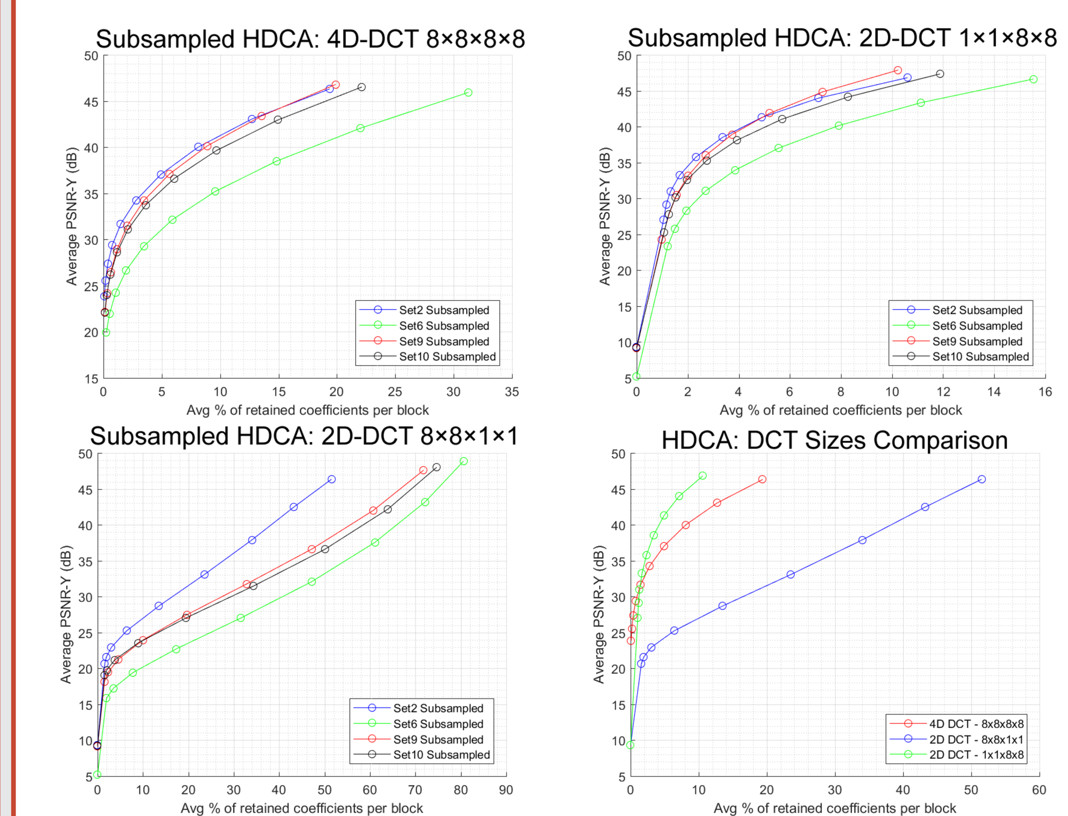
- Larger sparsity for 4D-DCT. It is worthy to exploit the 4D redundancy in a DCT-based coding scheme;
  - The exploration of 4D redundancy as a whole may lead to better coding efficiency;
- Inter-view sparsity is larger than the intra-view one: It is more effective to use an inter-view transform than an intra-view one;
  - 2D inter-view DCT is much closer to 4D DCT than 2D intra-view DCT;
- Similar behaviour for all lenslet-based datasets.

7 - Results: HDCA



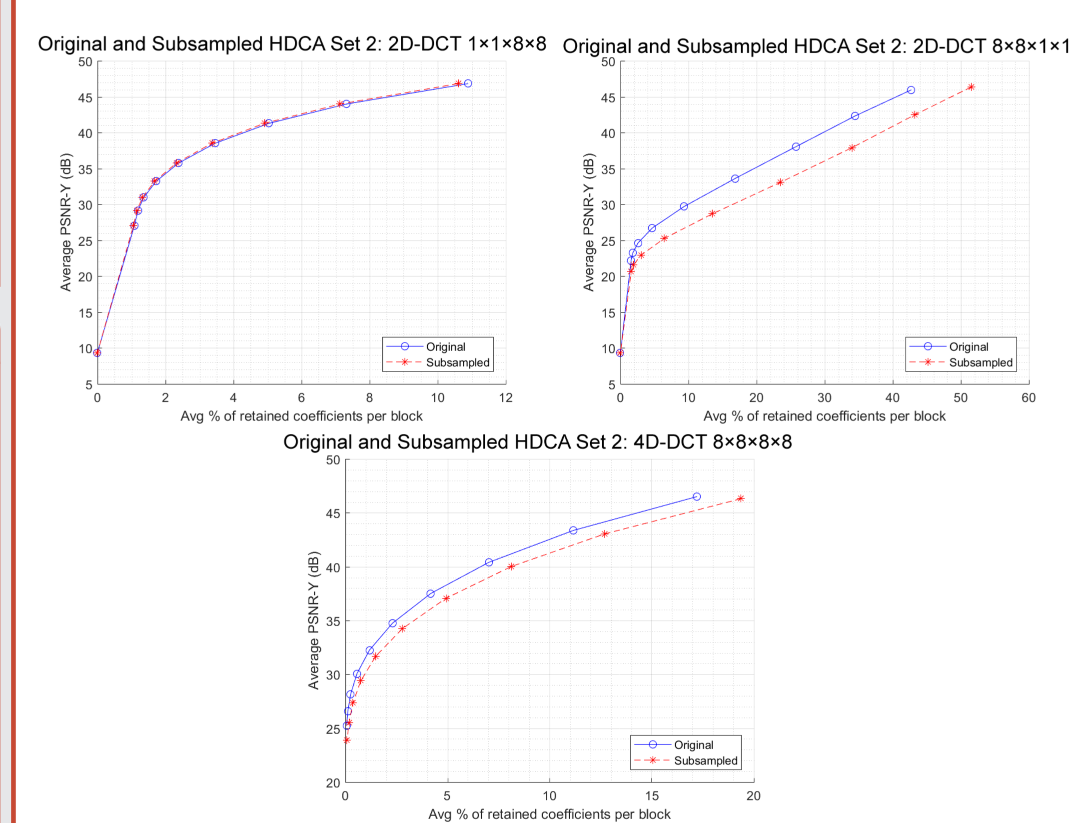
- The sparsity is larger for the 4D-DCT transform only for smaller percentage of retained coefficients;
  - 2D intra-view DCT has larger sparsity for larger percentage of retained coefficients;
- The sparsity is dominated by the intra-view redundancy, unlike the lenslet-based datasets;
- The inter-view sparsity is much smaller than the intra-view one;
- The HDCA dataset has much less 4D redundancy than the lenslet-based datasets.
- The same behaviour is observed for all HDCA datasets.

8 - Results: Subsampled HDCA



- Behaviour very similar to original HDCA;
- Sparsity is also dominated by the intra-view redundancy, with a much smaller inter-view redundancy.

9 - Comparison HDCA: Original vs Subsampled

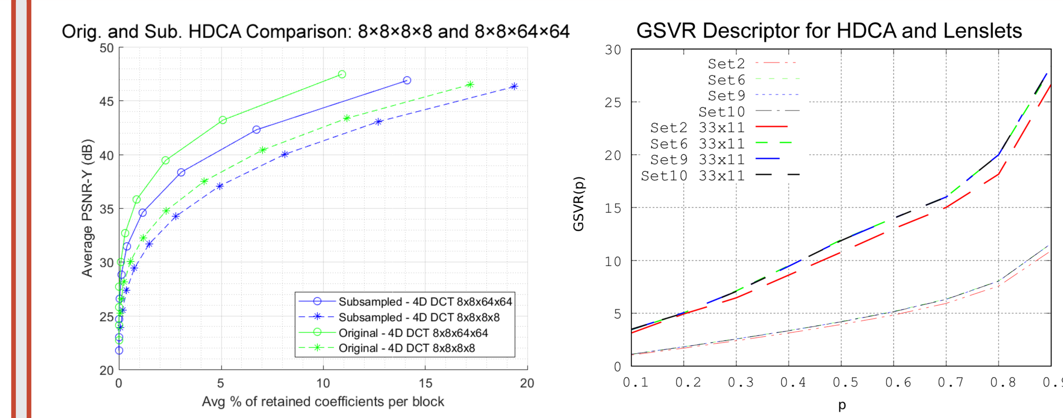


- Almost no difference between the original and subsampled datasets for 2D intra-view DCT - as expected;
- Large difference for 2D inter-view DCT: expected due to the larger spacing between adjacent views, introduced by the view subsampling;
- The difference between the original and subsampled HDCA datasets in terms of 4D redundancy is quite large.

10 - Comparison: HDCA vs Lenslets

- Lenslet-based datasets have 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;
- HDCA datasets have a much smaller amount of 4D redundancy than the lenslet-based datasets;
- It is likely that the coding solutions that are more efficient to the lenslet-based datasets will not be the more efficient ones to the HDCA datasets.

11 - HDCA: DCT $8 \times 8 \times 8 \times 8$ vs DCT $8 \times 8 \times 64 \times 64$



- DCT sizes differ only on the intra-view dimensions (8×8 and 64×64);
- The datasets differ only in the inter-view redundancy;
- For the same sparsity level, the difference in PSNR values between the two 4D-DCT sizes is higher for the original datasets;
- Conclusion: intra-view block size impacts on the exploitation of the inter-view redundancy. Larger intra-view dimensions are better.
  - Note: if intra-view dimensions are too large, the intra-view redundancy cannot be well exploited.
- The results are in accordance with the GSVR curves.

12 - Final Remarks

- Lenslet-based and HDCA datasets have a great amount of 4D redundancy that can be explored for coding purposes;
- Not exploiting the 4D redundancy as a whole may be a limitation to the design of LF codecs;
- HDCA and lenslet-based datasets may require distinct coding solutions due to the different nature of their 4D redundancy;
- The conclusions are restricted to the JPEG Pleno datasets. A more extensive study should be done using more general LF data.

Acknowledgements

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

References

[1] Adelson, E. H. and Bergen, J. R., "The Plenoptic Function and the Elements of Early Vision", Computational Models of Visual Processing, MIT Press, pp. 3–20, 1991.
[2] Ng, R., Digital Light Field Photography, Ph.D. Thesis, Stanford University, Stanford, USA, 2006.
[3] Dansereau, D.G., Plenoptic Signal Processing for Robust Vision in Field Robotics, Ph.D. Thesis, University of Sydney, Sydney, Australia, 2014.
[4] Johannsen O., Sulc A, et al, "What Sparse Light Field Coding Reveals about Scene Structure,"IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, June 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, Canada, September 2015.
[6] Shi, L. et al, "Light Field Reconstruction Using Sparsity in the Continuous Fourier Domain," ACM Transactions on Graphics, ACM, pp. 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," 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, June 2016.
[12] ISO/IEC JTC 1/SC 29/WG1N77016, "Core Experiments Set 2 for JPEG Pleno,"Macau, China, October 2017.
[13] Pereira, M. P., et al, "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.