

“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Joint Exploration Model based Light Field Image Coding: A Comparative Study

Huy Phi Cong^{1,2}, Stuart Perry³, Trinh Anh Vu¹, Xiem HoangVan¹

¹ VNU-University of Engineering and Technology

² Posts and Telecommunications Institute of Technology

³ University of Technology Sydney

huyphc@ptit.edu.vn, stuart.perry@uts.edu.au, vuta@vnu.edu.vn, xiemhoang@vnu.edu.vn

Abstract - The recent light field imaging technology has been attracting a lot of interests due to its potential applications in a large number of areas including Virtual Reality, Augmented Reality (VR/AR), Teleconferencing, and E-learning. Light Field (LF) data is able to provide rich visual information such as scene rendering with changes in depth of field, viewpoint, and focal length. However, Light Field data usually associates to a critical problem - the massive data. Therefore, compressing LF data is one of the main challenges in LF research. In this context, we present in this paper a comparative study for compressing LF data with not only the widely used image/video coding standards, such as JPEG-2000, H.264/AVC, HEVC and Google/VP9 but also with the most recent image/video coding solution, the Joint Exploration Model. In addition, this paper also proposes a LF image coding flow, which can be used as a benchmark for future LF compression evaluation. Finally, the compression efficiency of these coding solutions is thoroughly compared throughout a rich set of test conditions.

Keywords: *light field, plenoptic, super-pixel, coding standards, compression efficiency, complexity.*

I. INTRODUCTION

Light Field (LF) refers to the capture of information on the angle of incidence of light rays on an image sensor in addition to the spatial and intensity information traditionally captured. The capture of angular information provides data with rich information. Early work on this topic included the image-based rendering (IBR) system proposed by Levoy and Hanrahan in 1996 [1]. During the last two decades, there has been a huge improvement in LF techniques, e.g. acquisition, rendering, and sampling techniques. In particular, many cameras have been developed to capture LF data, for instance, Lytro LF, Illum [2] and Raytrix [3]. These cameras offer the amazing features of LF data such as changing perspective and viewpoints, digital refocusing, three-dimensional (3-D) data extraction, and depth estimation [4].

The massive data associated with LF technology brings about the amazing features listed above but also comes with disadvantages in the terms of data storing and transmission. In particular, this type of data presents sampling problems with

irregular sample spacing and the need for spatio-directional sampling kernels as well as the high dimension data presenting a challenge for classical preprocessing tasks. However, with its rich information, LF data has a promising future. LF data has the potential to enhance post production for movie processing, enabling new immersive experiences in mixed reality, teleconferencing, and improving visual content in virtual and augmented reality (VR-AR) [5]. In addition, LF displays allow for both eye vergence and accommodation in the display of 3D data mitigating one of the major causes of “VR sickness” [6].

Toward this goal, various techniques have been proposed to manage the performance compression and assessment methodology. The proposed techniques mostly focus on the compression multi-views and multi-focus from sub-aperture images derived from LF data [7]. A recent improvement is a self-similarity compensated prediction based HEVC coding solution for LF [8]. The key idea is to use bi-prediction of the blocks in a search window with similar information from the LF image. Another interesting idea is that of Dong Liu, *et al.* to compress all constitute viewpoints of a LF image as a pseudo sequence. This approach achieved significant coding gain compared to image-based coding schemes [9]. The first standards benchmarking is provided in [10] to compare compression efficiency of LF data between several coding solutions including HEVC, H.264/AVC, JPEG2000 and JPEG. However, this evaluation does not take into account the most recent coding solution, Joint Exploration Model, currently being developed by a Joint Video Exploration Team (JVET) of ITU-T VCEG [11] and ISO/IEC MPEG [12]. Besides, the well-known Google video platform, the VP9 video coding software [13] does not carefully evaluated for LF data compression.

Therefore, this paper proposes a comprehensive analysis for compression efficiency of LF image coding with not only the available image/video coding standards, e.g., Google/VP9 [13], H.264/AVC [14], JPEG-2000 [15], and HEVC [16] but also with the most recent JEM solution. A performance analysis for each test case is provided in order to have a clear

view of the recent achievement of emerging LF coding in relation to standard datasets.

The remainder of this paper is organized as follows. Section 2 briefly describes the background work on LF imaging and relevant image/video coding standards while Section 3 proposes a JEM based LF image compression model. Afterwards, Section 4 analyzes the experimental results for a rich set of test conditions. Finally, Section 5 gives some conclusions and future works.

II. BACKGROUND WORKS

A. Light Field Images

In this paper, a LF image is derived from the raw sensor output of a plenoptic camera [7]. Normally, the raw sensor output is not displayed. In order to convert and view LF data, a Light Field Toolbox (LFT) [17], were commonly used. The LF data can be captured by various cameras with different construction such as Raytrix, Lytro Illum, Stanford Multi-Camera array [18], etc., thus, each camera provides different data. The LF data used in this paper includes:

1) The data in [19] captured by Lytro Illum camera, it presents with resolution of 7728×5368 samples and with “GRBG” color order in a Bayer filter as shown in Fig. 1. After unpacking and converting with the LFT, a full LF image can be displayed as a 2D 7728×5368 image, displaying each of the lenslet images. In order to utilize the multi-views and multi-focal of the LF image, it has to transform 2D image to a 4D array which provides multi dimensions as $15 \times 15 \times 434 \times 625 \times 4$ where 15×15 represents the number of views, 434×625 represents the resolutions of each view, and 4 relates to color space of RGB and a weighting image component [7].

2) The data in [20] is set of various images with different view and focus which is captured by Lego Gantry system, designed and built by Andrew Adams from Computer Graphics Laboratory, Stanford university. This data is slightly different with data in [18], which is set of normal 2D RGB image with moderate resolution, e.g.: 1400×800 , 1024×640 , 1280×960 , etc. Besides, it has 289 views on a 17×17 grid and it is able to apply with Stanford software to view these Light Fields online through web browser. Hence, this data is possible to display without unpacking and converting. In this paper, it considers as massive lenslet as data in [19].

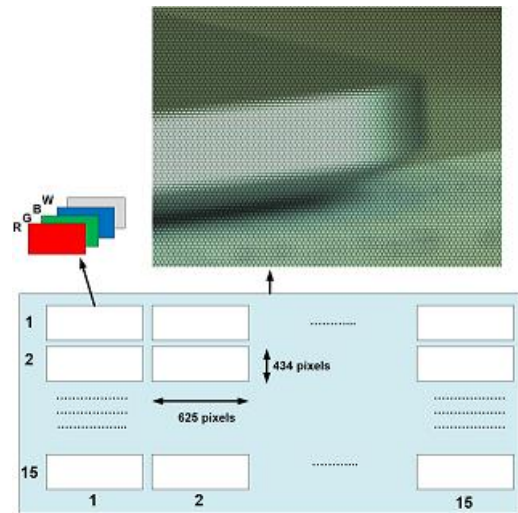


Fig 1. LF data structure as produced by LFT with massive lenslet [18]

B. Image/Video Coding Standards

LF data contains a rich set of information and can be presented as a still image. Hence, both image and video coding standards can be applied to compress LF data such as JPEG2000, H.264/AVC, HEVC, and Google/VP9. In this section, a brief description of these standards is presented.

1) JPEG-2000

JPEG-2000 is one of the latest image coding standards [15], which was created by the Joint Photographic Experts Group in order to cover the weaknesses of previous standard – JPEG [15] such as: poor bit-rate compression, large image handling, single compression architecture, etc. JPEG-2000 does not only provide higher compression efficiency, but also creates a new image representation with a rich set of features compared to JPEG, i.e., scalability and random access. The Wavelet transform is one of the main reasons for such interesting features.

2) H.264/AVC

H.264/AVC is a commonly used format for compression and distribution of video content which is developed by the ITU-T Video Coding Experts Group (VCEG) together with Moving Picture Experts Group (MPEG). It is a block-oriented video compression standard based on motion compensation [14]. There are two main layers in H.264/AVC, one is a video coding layer (VCL) which specifies motion compensation, transform coding, and entropy coding, while the other is a network abstract layer (NAL) that encloses coded slices into the network object in the network.

3) H.265/HEVC

High Efficiency Video Coding (HEVC), also known as H.265, is the latest video compression standard. As reported in [16] for a wide set of video test sequences, HEVC achieved a significant compression gain when compared to

the prior H.264/AVC standard. This comes from a large number of improvements in both coding structure, i.e. coding tree division with up to 64×64 , Merge coding mode, Sample adaptive offset, etc. [16] and prediction coding improvements, i.e., intra, inter predictions.

4) Google/VP9

VP9 is a next generation open source video codec from Google. It is competitive directly to the state-of-the-art video codec HEVC. Google/VP9 has several technologies in common with HEVC such as: quad tree (64×64) in prediction block size, Discrete Cosine Transform (DCT) in transform types, and prediction modes such as inter and intra. However, there are few noticeable differences, i.e. the block decomposition of VP9 is similar to a quad tree but has 4 partition modes: none, horizontal, vertical and split compare to two partition modes, none and split, of HEVC [13]. In addition, transform types of Google/VP9 are not only DCT, but also a modified Asymmetric Discrete Sine Transform (ADST) [13].

III. JEM BASED LIGHT FIELD IMAGE CODING

A. Overview of JEM Video Coding Solution

Although HEVC [16] has achieved a significant improvement in terms of compression efficiency when compared to the prior H.264/AVC standard [14], it is foreseeable that even better compression will be needed in the future, both in the context of traditional and new application domains. In this regard, a joint exploration activity has been started by ISO/IEC MPEG and ITU-T VCEG under the umbrella of their Joint Video Exploration Team (JVET). Such future standardization could take the form of additional extension(s) of HEVC or an entirely new standard [21, 22]. Example sources include camera-view content, screen content, consumer generated content, virtual reality/360° omnidirectional content, and high dynamic range content, while example applications include broadcast (with live or pre-authored content), real-time video conferencing, video chat, on-demand viewing, storage-based media replay, and surveillance with fixed or moving cameras.

B. JEM Enhanced Coding Techniques

The basic architecture of JEM is similar to the HEVC standard [21, 22], notably with a hybrid predictive and transform coding based approach. However, to achieve a further compression improvement, several advanced coding tools have been proposed for JEM architecture, notably: 1) Intra prediction improvement with 67 prediction modes, four-tap intra interpolation filter, and cross-component prediction; 2) Inter prediction improvement with sub-prediction unit based motion vector (MV) prediction, adaptive MV resolution, overlapped block motion compensation and bi-directional

optical flow; 3) Transform coding improvement with adaptive multiple core transforms, secondary transform, and signal dependent KLT [21, 22] transform and 4) Entropy coding improvement with the context model selection for transform coefficient levels, multiple adaptation rate probability estimation, and initialization for context models.

C. JEM based Light Field Image Coding

In the proposed LF coding solution, the JEM is used to compress LF data. The process of obtaining a particular view of the LF data can be demonstrated as in Fig. 2.

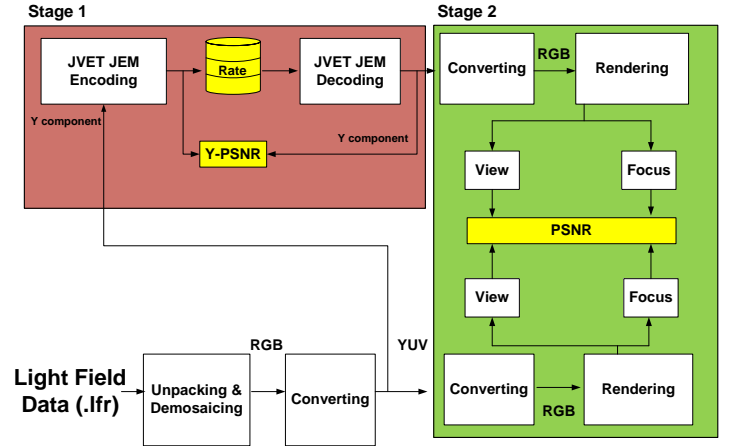


Fig 2. Light Field Coding Evaluation Framework

As shown in Fig. 2, it is able to get a specific view and focus of a conventional 2D image with two stages, primary stage for compressing LF data and secondary stage for displaying the LF data. According to high-dimensionality of the LF image, LFR input at stage one is chosen to unpack and de-mosaic. After providing the LFR output as an RGB and then transforms it to the YUV color space using an 8-bit unsigned integer representation in order to be compatible with available coding software, i.e. JPEG-2000, H.264/AVC, H.265/HEVC, Google/VP9 and JEM. In this work, the Y component of the demosaiced image is used to compute primary RD and the size of encoded stream becomes the rate for the following RD performance analysis.

IV. EXPERIMENTAL RESULTS

In this section, the various test cases are provided, and we examine the compression efficiency and complexity with various codecs.

A. Test Methodology

In this paper, the test methodology uses various coding standards to compute the primary bit rate and quality of the full LF image. The LF image with a resolution of 7728×5368

samples the input to the performance test. As usual, the luminance (Y-component) of image is an effective value to compression process rather than chromatic value [10]. Thus, raw LF image after being demosaiced, is converted to YUV format with only the Y value is for subsequent processing. Each coding standard will be set to default settings and used to compress the Y component of the 6 demosaiced LF images from [19] and 6 LF from the gantry [20] shown in Fig 3.

For content visualization, thumbnails of the selected LF images are shown in Fig. 3.

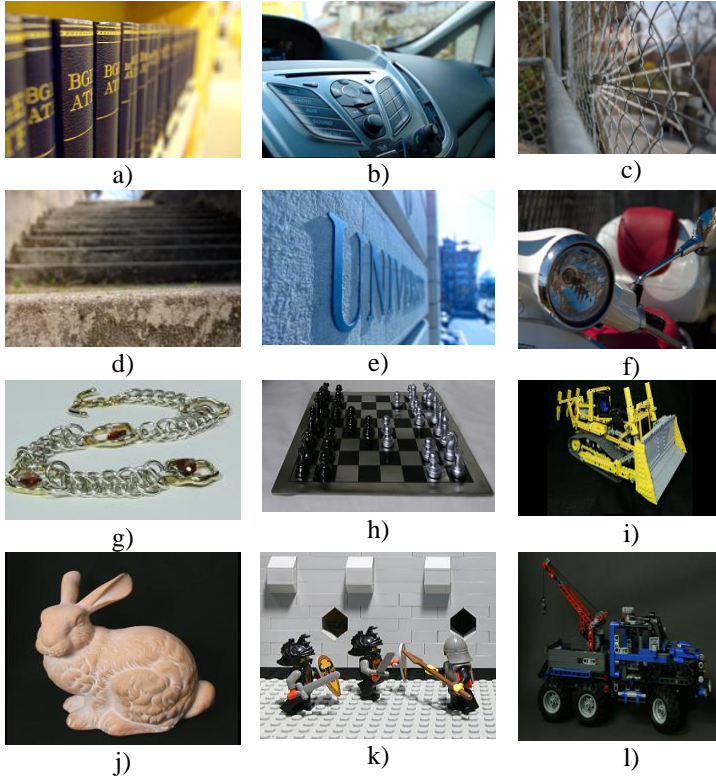


Fig 3. Thumbnails of Light field images in [15] from (a) to (f) and in [23] from (g) to (l) : (a) Books, (b) Car_Dashboard, (c) Spear_Fence_2, (d) Stairs, (e) University, (f) Vespa, (g) Bracelet , (h) Chess, (i) Lego Bulldozer, (j) The Stanford Bunny , (k) Lego Knights , (l) Lego Truck

The implementation of the JPEG-2000 encoding and decoding used in this paper was OpenJPEG version 2.1.2 [23]. This is an open-source JPEG 2000 codec and the software was set to default for all settings. The rate distortion (RD) points were set by a compression ratio (CR) ‘r’ in 4 ratios, i.e., 10:1, 20:1, 40:1, and 100:1.

A basic coding unit in H.264/AVC is the Macroblock that is encoded in either intra or inter mode. In this paper, H.264/AVC intra mode is adopted for compression performance as no temporal correlation exists within the LF image. In order to have a fair-comparison, the software used to encode and decode is FFMPEG version 3.3.2 [24] which

supports x264 encoding, an H.264 compatible video implementation. It is a very high quality encoder and produces remarkable quality bit streams. The quantization parameter is set at 19, 26, 31, and 38 corresponding to the compression ratio ‘r’ of OpenJPEG.

In the same manner as H.264/AVC, FFMPEG version 3.3.2 is used to encode and decode with x265 codec similar with x264 with default setting and quantization parameter as shown in Table I.

For the assessment, we use a codec supporting encode and decode of Google/VP9 (libvpx-vp9 version 1.6.1) which is included in FFMPEG version 3.3.2. The default configuration of Google/VP9 is to set “-b:v 0” along with a constant rate “crf”, and the “-deadline” parameter set. This configuration provides a good balance of quality and rate.

In order to have JEM compression performance evaluation, outstanding codec HEVC (x265) is chosen with data in [20] for particular test case with various resolution. The default setting stays still with JEM and HEVC (x265), except the quantization parameter slightly changing, i.e. JEM encodes at qp 22, 27, 32, 37 while HEVC (x265) encodes at qp 28, 33, 38, 42.

TABLE I. QUANTIZATION PARAMETERS (QPS) USED WITH COMPRESSION RATIO FOR ALL CODING SCHEMES

Codec	R1	R2	R3	R4
JPEG-2000	10:1	20:1	40:1	100:1
x264	19	26	31	38
x265	18	25	30	37
VP9	17	32	41	53

B. Standards Compression Performance Evaluation

For video content, HEVC (x265) and Goolge/VP9 (libvpx-vp9) naturally are expected to outperform other codecs such as the H.264/AVC (x264) and JPEG-2000. In this evaluation, the RD performance is compared and presented in Fig. 4 and the Bjøntegaard Delta (BD)-Rate [25] saving compared to the JPEG-2000 is computed in Table II.

TABLE II. BD RATE [%] SAVING COMPARED TO JPEG-2000

LF images	H.264/AVC	VP9	HEVC
Books	4.43	-19.17	-14.01
Car_Dashboard	4.98	-19.55	-19.65
Spear_Fence_2	0.11	-23.46	-25.52
Stairs	-1.46	-23.28	-24.51
University	3.05	-20.54	-21.16
Vespa	2.13	-31.09	-29.03
Average	2.21	-22.85	-22.31

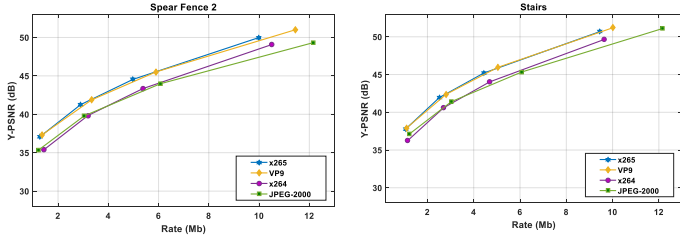


Fig 4. RD performance comparison of light field test images Spear Fence 2 and Stairs with HEVC standard

From the BD rate assessment of all standards with JPEG-2000, as shown in Table II and Fig. 4, some conclusions can be drawn:

- When compressing a LF picture, only spatial correlation is exploited, the compression efficiency of the H.264/AVC video coding standard is close to the JPEG image coding standard, notably with only about 2.2% BD rate increase while providing a similar perceptual quality.
- HEVC (x265) is still the best choice for compression efficiency for all situations with the saving in bitrate approximately 22% compare with JPEG-2000. This mainly comes from a large number of Intra prediction modes used in HEVC [16].
- The results for Google/VP9 (libvpx-vp9) indicate that compression efficiency is close to that of HEVC (x265), notably in intra mode (one frame compression).

C. JEM Compression Performance Evaluation

From 6 particular tests with different resolutions of Light Field data in [20], JEM has shown outperformance with HEVC (x265), it is true to be a future coding.

TABLE III. JEM VS. HEVC FOR LF IMAGE CODING

LF images	BD-Rate	BD-PSNR
Bracelet	-37.31	3.66
Lego Bulldozer	-41.13	2.84
The Stanford Bunny	-49.93	2.51
Chess	-44.90	3.25
Lego Knights	-39.78	3.30
Lego Truck	-45.64	2.65
Average	-43.12	3.04

Following the result as shown in Table III and Fig. 5, it can sum up as below:

- JEM is not only inherited all the advance technology from HEVC, but also upgraded to several advantages to new trend, hence, JEM presents an impressive compression efficiency with saving in bitrate about 43% compare to HEVC(x265) in almost test cases while still providing a similar perceptual quality.
- Beside bitrate saving, quality of JEM is outperformance with HEVC(x265) by increasing around 3dB in average.

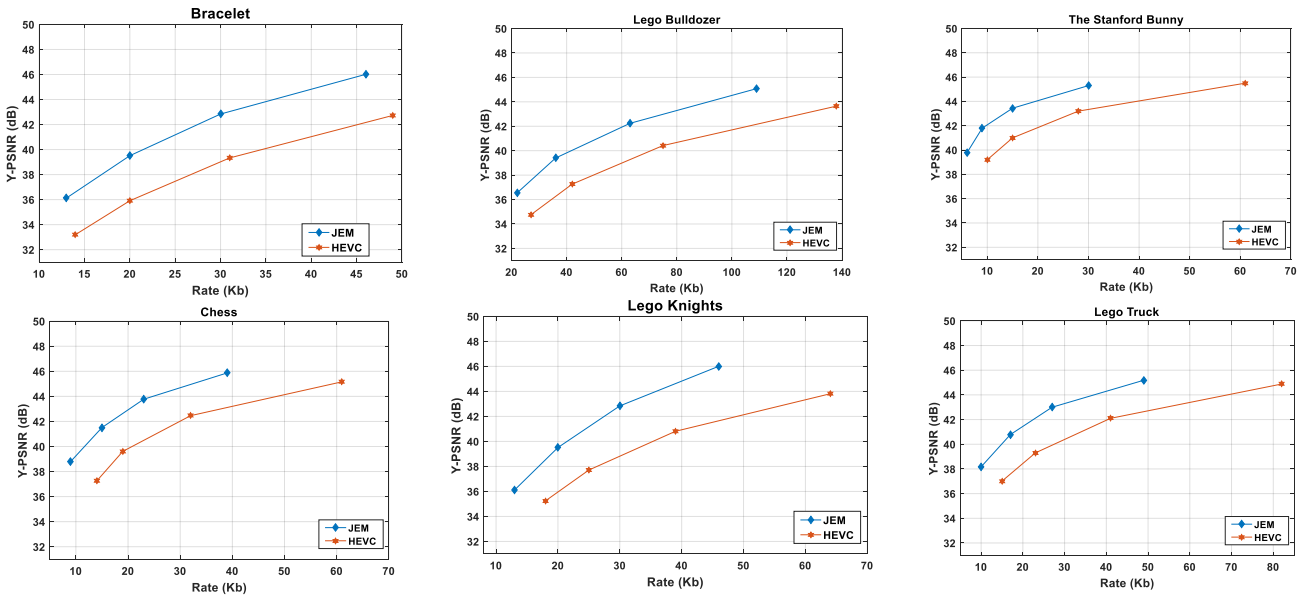


Fig 5. RD performance comparison of JEM and HEVC with 6 light field test images from [20]

V. CONCLUSIONS

This paper proposes a comprehensive performance evaluation for LF image coding with the popular advanced image/video coding standards as well as a recent video coding solution. As shown, the compression performance of the HEVC (x265) and Google/VP9 (libvpx-vp9) significantly outperforms JPEG-2000 while the compression performance of the H.264/AVC (x264) is close to that of JPEG-2000. Especially, JEM, a recently developing video coding solution, has shown an impressive performance for both bitrate saving and quality improvement comparing with the well-known HEVC standard. Therefore, it is suggested to deploy JEM for compressing LF data. Future work may further analyze the trade-off between the compression efficiency and complexity associated to the JEM based of LF image coding proposal.

ACKNOWLEDGEMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01- 2016.15.

REFERENCES

- [1] Marc Levoy and Pat Hanrahan, "Light field rendering," in Proc. SIGGRAPH, 1996, pp. 31–42.
- [2] Lytro camera, <https://www.lytro.com/>
- [3] Raytrix, <https://www.raytrix.de/>
- [4] Ivo Ihrke, John Restrepo, and Loïs Mignard-Debise, "Principles of Light Field Imaging", IEEE Signal Processing Magazine, 2016
- [5] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, Peter Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," IEEE Multimedia, Oct-Dec 2016.
- [6] M. Lambooi and W. Ijsselstein, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review", Journal of Imaging Science and Technology, 53(3), 030201-030201-14, 2009.
- [7] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, Pat Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," Stanford Tech Report CTSR, 2005.
- [8] C. Conti, P. Nunes and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016, pp. 1-4.
- [9] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, Wenjun Zeng, "Pseudo-Sequence-Based Light Field Image Compression", Multimedia & Expo Workshops (ICMEW), IEEE International Conference, July 2016
- [10] G. Alves, F. Pereira and E. A. B. da Silva, "Light field imaging coding: Performance assessment methodology and standards benchmarking," 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016, pp. 1-6.
- [11] ITU-T VCEG : <http://www.itu.int/en/ITU-T/studygroups/com16/video/Pages/default.aspx>
- [12] ISO/IEC MPEG: <https://mpeg.chiariglione.org/>
- [13] D. Mukherjee *et al.*, "A Technical Overview of VP9 – The Latest Open-Source Video Codec," SMPTE 2013 Annual Technical Conference & Exhibition, Hollywood, CA, USA, , pp. 1-17, 2013.
- [14] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.
- [15] M. Rabbani, R. Joshi, "An overview of the JPEG2000 still image compression standard", Signal Processing: Image Communication 17, 3–48, 2002.
- [16] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
- [17] D. Dansereau, "Light Field Toolbox v0.4," <http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4>.
- [18] Stanford Multi-Camera Array, <http://graphics.stanford.edu/projects/array/>
- [19] M. Režábek and T. Ebrahimi, "New Light Field Image Dataset," 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016.
- [20] Andrew Adams, "Light Fields from the Lego Gantry", <http://lightfield.stanford.edu/lfs.html>
- [21] Gary Sullivan, "Video Coding: Recent Developments for HEVC and Future Trends", Data Compression Conference, Snowbird, Utah, March, 2016.
- [22] "Algorithm Description of Joint Exploration Test Model 6", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1001-v3.
- [23] OpenJPEG software, <http://www.openjpeg.org/>
- [24] FFMPEG software, <https://www.ffmpeg.org/>
- [25] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", Doc. ITU-T SG16 VCEG-M33, Austin, TX, USA, April 2001.