Impact of Light Field Compression on Focus Stack and Extended Focus Images

Mira Rizkallah*, Thomas Maugey†, Charles Yaacoub‡ and Christine Guillemot†
*IRISA/Université Rennes 1
†INRIA Rennes Bretagne-Atlantique
‡Holy Spirit University Of Kaslik (USEK)

***Abstract*—Light Fields capturing all light rays at every point in space and in all directions contain very rich information about the scene. This rich description of the scene enables advanced image creation capabilities, such as re-focusing or extended depth of field from a single capture. But, it yields a very high volume of data which needs compression. This paper studies the impact of Light Fields compression on two key functionalities: refocusing and extended focus. The sub-aperture images forming the Light Field are compressed as a video sequence with HEVC. A focus stack and the scene depth map are computed from the compressed light field and are used to render an image with an extended depth of field (called the extended focus image). It has been first observed that the Light Field could be compressed with a factor up to 700 without significantly affecting the visual quality of both refocused and extended focus images. To further analyze the compression effect, a dedicated quality evaluation method based on contrast and gradient measurements is considered to differentiate the natural geometrical blur from the blur resulting from compression. As a second part of the experiments, it is shown that the texture distortion of the in-focus regions in the focus stacks is the main cause of the quality degradation in the extended focus and that the depth errors do not impact the extended focus quality unless the light field is significantly distorted with a compression ratio of around 2000:1.**

I. Introduction

During the last two decades, there has been a growing interest in Light Fields. Many acquisition and sampling techniques were envisioned in order to capture the light information present in a scene using arrays of cameras, plenoptic cameras or moving cameras [1], [2], [3], [4]. Essentially, a captured light field comprises geometrical information of the scene along with texture information. The light field representation enables various applications such as digital refocusing, depth estimation, changing perspective and viewpoints, simulating captures with different depth of fields and 3D reconstructions.

This comes at the expense of collecting large volumes of high-dimensional data, which appears to be the key downside of light fields. For example, a modest four dimensional light field, captured by a plenoptic camera comprising a 256x256 array of microlenses with 32x32 photosensors behind each microlens, yields a storage footprint of around 200 Mbytes, which is significantly large for a photograph. Therefore, to realize practical applications of light fields, it is essential to efficiently compress this data without compromising the ultimate quality and more importantly, without leaving an undesirable effect on the targeted post-capture processing.

Some research effort has been dedicated in the past years to the design of light fields compression schemes based on vector quantization [5], transform coding [6] [7], statistical representations [8], multiview video compression and disparity compensation techniques [9], or adaptations of HEVC [10], [11], [12]. However, very little attention has been given to the impact of compression on the image creation functionalities.

In this paper, we aim at analyzing the impact of compression on two post-capture image rendering functionalities: refocusing and extended focus. The extended focus image is characterized by the fact that all its pixels are in-focus whereas the focus stack images have in-focus and out-of-focus pixels in different regions of the scene.

Toward this goal, we first compress the light field using HEVC, considering the set of sub-aperture images as a video sequence. The focus stack images are then computed by shifting and adding the compressed sub-aperture images. Afterwards, the scene depth map is estimated from the compressed sub-aperture images and the computed focus stack using the method described in [13]. The observed quality of the focus stack and the extended focus images show that one can vary the QP parameter up to 32, and decrease the light field data volume from 50 Mbytes down to 75 kbytes without visually impacting the quality of the rendered images.

Looking at a refocused image after compression, the blur due to compression is unnoticed in out-of focus regions where the quantization blur is mixed to natural geometry blur. As for the extended focus, the compression blur is visible in the entire image. Comparing with the refocused and extended focus images computed from the original light field, it appears that the refocused images are visually more robust to compression than the extended focus image. Traditional metrics such as PSNR fail to accurately reflect this unbalanced robustness. For this reason, a dedicated metric naturally differentiating the geometrical blur present in out-of-focus regions from the blur introduced by compression is then considered based on contrast and gradient measurements. Furthermore, it is shown that the depth map used for creating the extended focus image does not impact the extended focus even for a compression factor of 2000, and that the extended focus image quality degradation essentially results from the texture distortion of the in-focus pixels of the focus stack images.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

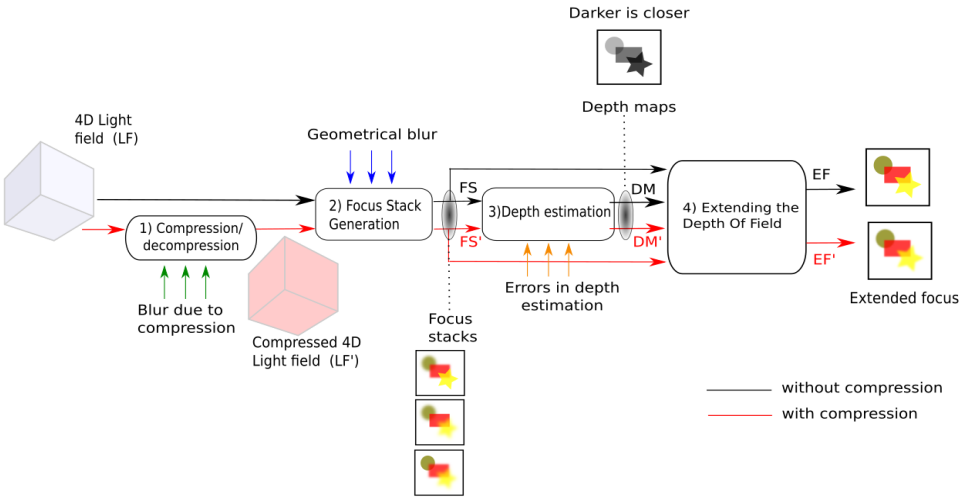


Fig. 1: Processing chain considered in this work with four main stages: 1) Compression, 2) Focus stack generation, 3) Depth estimation and 4) Extending the depth of field.

II. Light Fields processing chain

The light field processing chain considered in this work is depicted in the scheme of Fig. 1. It proceeds in four phases: compression, generating the focus stack, depth estimation and finally extending the depth of field. We explain each of these steps separately in the following.

The light field sub-aperture images are first assembled, following a spiral scan line (see Fig. 2), as a video sequence which is coded with HEVC. The spiral scan line is justified by the fact that the luminosity of the sub-aperture images varies progressively going from the center outwards due to the spherical shape of the camera main lens. The method is compared to previous multiview and disparity compensation techniques [9] in Fig. 3 for the Buddha light field[1]. Given the significant improvement in the rate-distortion efficiency, this method is used in our analysis in the following sections.

Once the light field has been compressed, 32 photographs focused at different depths, forming the so-called focus stack, are computed by shifting and adding the sub-aperture images [1] [2] as in Eq. (1)

$$E_{\alpha_i}(x,y) = \frac{1}{(\alpha_i F)^2} \sum_u \sum_v L^{(u,v)}(X_s + x, Y_s + y) \qquad (1)$$

$$X_s = u(1 - \frac{1}{\alpha_i}); Y_s = v(1 - \frac{1}{\alpha_i}),$$

where $E_{\alpha_i}(x,y)$ denotes the pixel value at position (x,y) in the i^{th} refocused image E_{α_i} at depth α_i. The value F represents the distance between the lens and the reference film planes. Focusing at different depths α_i corresponds to changing the separation between the lens and the film plane, hence the multiplication $\alpha_i F$.

In the experiments reported in the paper, the parameter α_i takes 32 different values in the interval [0.6, 1.8]. This parameter actually controls the position of the focus plane.

[1]buddha4.tar.gz at http://graphics.stanford.edu/software/lightpack/lifs.html

$L^{(u,v)}$ is the sub-aperture image from the (u,v) position on the main lens aperture. X_s and Y_s are the shift amounts.

The image creation process is equivalent to shearing the 4D light field varying the slope of epipolar lines. At that stage, a natural geometrical blur appears in out-of-focus regions.

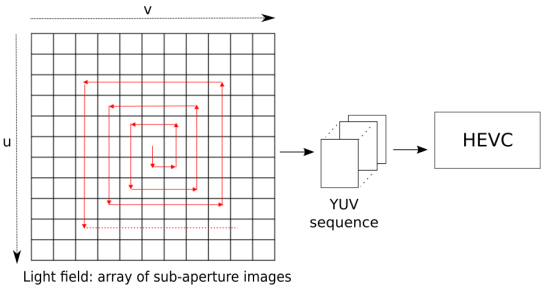


Fig. 2: Light field coding using HEVC: The sub-aperture images are coded in a video sequence following a spiral scan line.

Afterwards, the scene depth map is also estimated from compressed sub-aperture images and the computed focus stack using the method detailed in [13] which combines both defocus and correspondence cues. For each pixel, we seek the optimal contrast within a patch along refocused images thus the highest defocus response, and the minimum correspondence response minimizing the angular variance along the sheared light fields. The output of this phase is a depth map with each pixel value pointing out to one image of the focus stack. Once the depth map and the focus stack have been computed, the depth of field can be extended, focusing on all the scene at once leading to an image with no geometrical blur. The all in-focus image $E(x,y)$ is constructed from the depth map and the focus stack images as follows. For each pixel at position (x,y) and its corresponding depth α_i, the extended focus image $E(x,y)$ is formed by taking the pixel at position (x,y) in

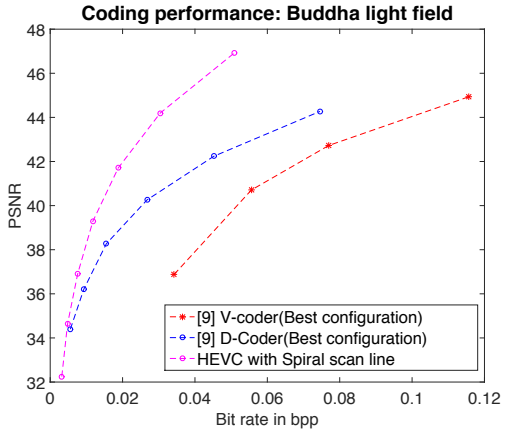


Fig. 3: Comparison of the coding efficiency of HEVC-Spiral (using QP =24, 28, 32, 36, 40, 44, 48) with the methods in [9] based on disparity compensated multiview coding techniques.

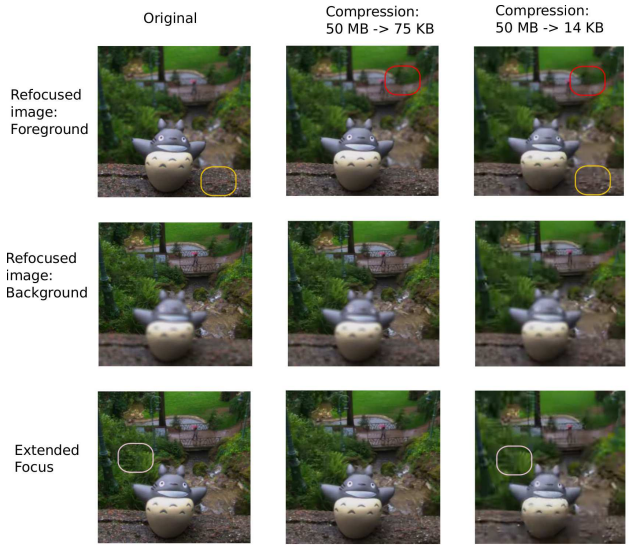


Fig. 4: Evolution of the quality of the refocused and extended focus images' while varying the compression ratio. The red patches refer to out-of focus regions of a refocused image, the orange ones to the in-focus regions without and with light field compression and the white ones to a region in the extended focus without and with light field compression.

the focus stack image $E_{\alpha_i}(x, y)$ ($E(x, y) = E_{\alpha_i}(x, y)$).

III. Subjective results and analysis

As depicted in Fig. 1, the four stages involve two types of blur: a geometrical blur caused by the refocusing procedure which is, by nature, good for image quality, and a quantization blur which decreases the perceived quality due to texture compression. Besides, some errors may occur in the depth estimation process, where a wrong depth is assigned to some of the pixels.

The added amount of blur can be observed in Fig. 4 where results of the Refocusing and Extended Focusing algorithms before and after compression of a natural Light Field are shown. The two types of blur are apparent in the refocused image. On the one hand, the red patch refers to an out-of-focus region in the background where the geometrical blur is mixed with compression blur after compression. In these regions, the compression blur is not visually perceptible. On the other hand, the orange ones refer to an in-focus region in the foreground only affected by the compression blur that thus becomes visible at high compression ratios. When combining the in-focus regions of all the refocused images in an extended focus image, only the blur due to compression remains such as in the white patches. Experimentally, we observed that we can achieve high compression ratios, attaining a range of 500:1 to 700:1 without altering the visual quality of the extended focus image. Afterwards, for very high compression ratios, the quality drops significantly leading to a fully blurred image.

We remark that the perceived quality of the refocused images is naturally higher compared to the extended focus. This is due to the fact that some parts of the refocused images are naturally blurred, and the probability of adding blur to in-focus regions is relatively small compared to the extended focus. Moreover, the extended focus is built from the focus stack, therefore the compression blur propagates from in-focus regions of the focus stack to the extended focus. However, this subjective quality trend does not exactly align with primary PSNR measurements depicted in Fig. 5. More precisely, we plot the PSNR of the rendered images as a function of the PSNR of the compressed Light field and its size after compression while varying the QP from 24 to 44. A decrease of 1 dB of the light field's PSNR leads to a decrease of 2 dB in terms of PSNR for both functionalities. In other words, the PSNR does not reflect the fact that the extended focus is more affected by compression than the focus stack[2]. This is mainly due to the fact that the PSNR does not differentiate the geometry blur from the compression blur.

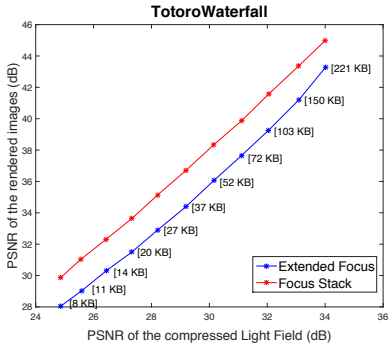


Fig. 5: The evolution of the refocused and extended focus images' quality as a function of the compressed Light Field quality.

For this reason, we propose in section IV, a dedicated metric based on gradient and contrast measurements that evaluates the amount of compression blur added to in-focus regions. With

[2]There, we are interested by the PSNR evolution rather than the PSNR value since the reference from which the MSE is computed is not the same.

this metric, we study the impact of compression on the amount of blur in the focus stack and the extended focus images. Afterwards, we examine the major cause of the extended focus quality degradation: errors in depth estimation or texture compression.

IV. Proposed metric

As explained in the previous section, the proposed metric aims at measuring the amount of compression blur added in the regions that are supposed to be in-focus in the focus stack and extended focus images. Let n_i and n_i' denote the numbers of in-focus pixels in a refocused image E_{α_i} (at depth α_i), before and after compression of the light field respectively. The percentage of pixels that become blurred after compression is expressed by:

$$\psi = \frac{n_i - n_i'}{N} \times 100 \tag{2}$$

where N denotes the total number of pixels in the evaluated image. n_i' is estimated from the difference in the gradient response of in-focus pixels before and after compression. The gradient responses are evaluated using the following gradient operator:

$$G(x,y) = \frac{1}{|W_D|} \sum_{(x',y') \in W_D} \Delta E_{\alpha_i}(x',y') \tag{3}$$

where E_{α_i} is the evaluated refocused image in the focus stack. Considering only pixels at positions (x,y) which were in-focus before compression, $G(x,y)$ is the gradient response averaged within a patch to improve robustness. W_D represents the window around the current pixel (x,y) with its size $|W_D|$ and Δ stands for the spatial gradient operator.

The gradient's variation is compared to a predefined threshold value T, to determine whether the pixel is still in-focus or not after compression.

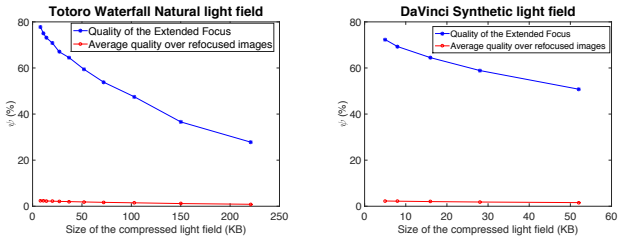
The calculated ratio ψ is averaged over the entire focus stack. For the extended focus image, ψ is computed under the perspective that everything is supposed to be in-focus (n_i is thus equal to N).

V. Experiments

In our experiments, HEVC Test Model (HM) reference software[3] was used to code the sub-aperture images. A GOP size of 4 was picked with a 'IBBBP' encoding scheme. Only the central viewpoint image is intra-coded. The QP was varied from 24 to 44. PSNR, SSIM and ψ are measured taking as reference the refocused and extended focus images computed from the original uncompressed light field.

A. Quality Evaluation: Focus Stack vs. Extended Focus

We used the ψ (Eq. (2)) to measure the percentage of blur added in the refocused and extended focus images after compression of the light field. For the experiments, we use two light fields: a natural one captured by a plenoptic camera and a synthetic one. *Totoro Waterfall* (available at [14]) is captured by a Lytro plenoptic camera and consists of 11x11 subaperture images containing 379x379 RGB pixels each. *Da Vinci* (available at [15]) is a synthetic light field comprising 9x9 Multiview images with 768x768 pixels each. Fixing T to 0.003 and the window radius to 3, ψ is plotted in Fig. 6 as a function of the size of the compressed light field. It shows the difference in the amount of blur added due to compression between the refocused and extended focus images.



(a) (b)

Fig. 6: Percentage of blurred in-focus pixels due to compression (ψ) for both the extended focus and refocused images. (a) Natural light field (Original size = 50 MB), (b) Synthetic light field (Original size = 136 MB)

After compressing both light fields, an average maximum of 3% of initially in-focus pixels become out-of focus in each refocused image. This is a very small amount compared to the extended focus image where a very high percentage of pixels (up to 80% at high compression ratios) becomes blurred due to compression.

The probability of adding blur in the in-focus regions is relatively small which makes the focus stack visually more robust to compression than the extended focus.

B. Impact of the Texture and Depth Estimation on the Extended Focus Quality

There are two possible causes of quality degradation of the extended focus image resulting from compression of the light field: depth map errors and focus stack texture distortion. For the purpose of investigating the impact of those two types of errors, we evaluate the quality of the extended focus image under different conditions. More precisely, three combinations of inputs to the extended focus estimation algorithm are tested separately: an original depth map DM with the focus stacks after compression FS'; a reconstructed depth map after compression DM' with original focus stacks FS and finally DM' with FS' which is the natural combination. We then plot, for each one of them, the quality evolution as a function of the size of the compressed light field. The quality is estimated with PSNR (Fig. 7 (a)(d)), SSIM (Fig. 7 (b)(e)), and ψ (Fig. 7 (c)(f)).

We see that using an original focus stack FS, the depth map DM'can be estimated from a natural light field compressed with a factor 3600 or a synthetic one with a factor of 8000 without altering the quality of the extended focus. On the other side, no additional gain in quality is observed with an original depth map DM and a reconstructed focus stack FS'. This

[3]Reference software for ITU-T H.265 high efficiency video coding Version 10/14 available at http://www.itu.int/rec/T-REC-H.265.2

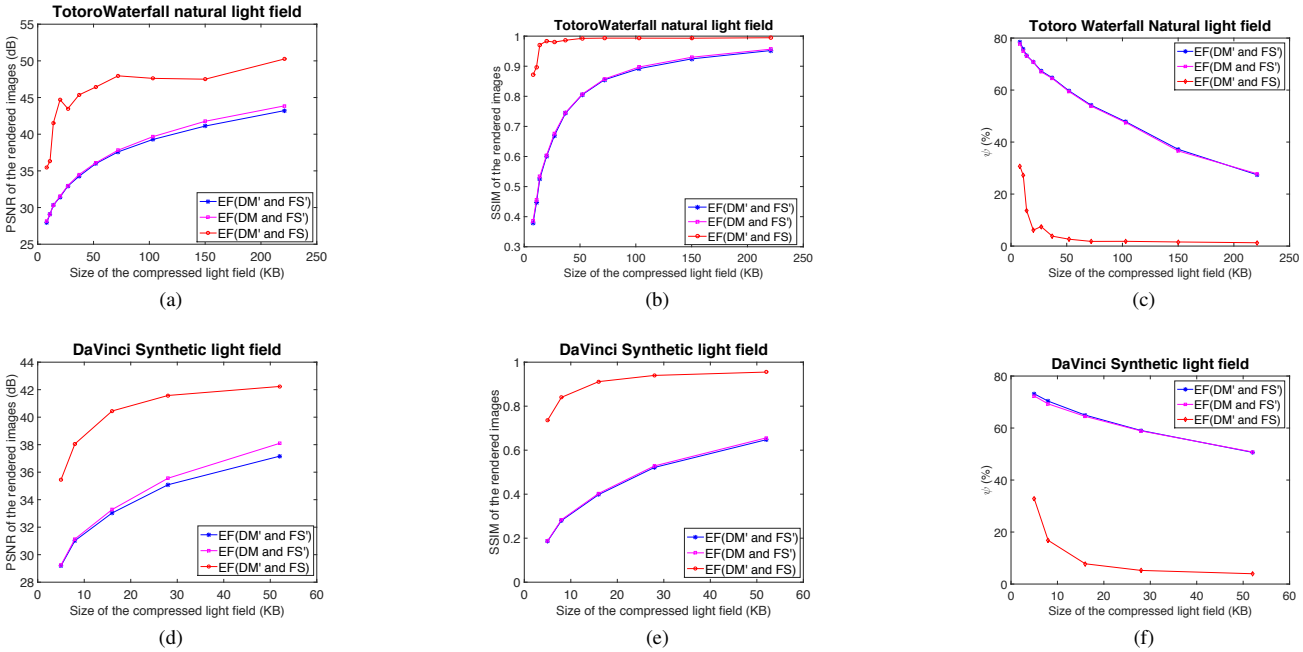


Fig. 7: Evolution of the quality (PSNR, SSIM and the dedicated metric ψ) of the Extended Focus (EF) with different input combinations: An original Depth Map (DM) and a reconstructed Focus Stack (FS'); an original Focus Stack (FS) and a reconstructed Depth Map (DM'); DM' and FS'. (a)(b)(c): A natural light field (Original size = 50 MB). (d)(e)(f): A synthetic light field (Original size = 136 MB).

shows that the extended focus quality degradation essentially results from the focus stack texture distortion.

VI. Conclusion

In this paper, we analyze the impact of light field compression on the quality of the focus stack and the extended focus images. It has been observed that a light field can be compressed by a factor of around 700 without altering the visual quality of both considered functionalities. Based on a dedicated quality metric adapted to the problem, we showed the focus stack is more robust to compression than the extended focus since already some parts of it are blurred natively. It has been also shown that the major cause of quality degradation in the extended focus is the texture distortion of in-focus regions in a focus stack while the depth estimation errors do not have a significant impact on the rendering quality. This study might be used to develop new performing coding schemes for Light Fields taking into account the quality of targeted applications.

References

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42.

[2] R. Ng, "Light field photography," Ph.D. dissertation, Stanford University, 2006.

[3] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[4] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," pp. 78 730X–78 730X–13, 2011.

[5] A. C. Beers, M. Agrawala, and N. Chaddha, "Rendering from compressed textures," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 373–378.

[6] G. Miller, R. S., and D. Ponceleon, "Lazy decompression of surface light fields for precomputed global illumination," in *Rendering Techniques '98*, 1998, pp. 281–292.

[7] M. Magnor, A. Endmann, and B. Girod, "Progressive compression and rendering of light fields," in *Vision, Modelling and Visualization*, 2000, pp. 199– 203.

[8] D. Lelescu and F. Bossen, "Representation and coding of light field data," *Graph. Models*, vol. 66, no. 4, pp. 203–225, Jul. 2004.

[9] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 10, no. 3, pp. 338–343, Apr. 2000.

[10] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod, "Light field compression using disparity-compensated lifting and shape adaptation," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 793–806, April 2006.

[11] Y. Li, M. Sjstrm, R. Olsson, and U. Jennehag, "Efficient intra prediction scheme for light field image compression," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 539–543.

[12] Y. Li, M. Sjstrm, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 80–91, Jan 2016.

[13] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *2013 IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia, 2013, pp. 673–680.

[14] A. Mousnier, E. Vural, and C. Guillemot, "Partial light field tomographic reconstruction from a fixed-camera focal stack," https://www.irisa.fr/temics/demos/lightField/index.html.

[15] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," http://hci.iwr.uni-heidelberg.de/HCI/Research/LightField/lfbenchmark.php.