

LIGHT-FIELD IMAGE COMPRESSION BASED ON VARIATIONAL DISPARITY ESTIMATION AND MOTION-COMPENSATED WAVELET DECOMPOSITION

Trung-Hieu Tran¹, Yousef Baroud¹, Zhe Wang¹, Sven Simon¹, David Taubman²

¹ The University of Stuttgart, Germany

² The University of New South Wales, Australia

ABSTRACT

This paper presents a compression framework for light-field images. The main idea of our approach is exploiting the similarity across sub-aperture images extracted from light-field data to improve encoding performance. For this purpose we propose a variational optimisation approach to estimate the disparity map from light-field images and then apply it to a motion-compensated wavelet lifting scheme. Making use of JPEG2000 for coding all high-/low-pass sub-band views as well as disparity map, our approach can therefore support both lossless and lossy compression. The coding framework is tested with both synthetic and real-world light-field dataset. The experimental results demonstrate that our approach outperforms JPEG-LS and the direct application of JPEG2000 in both lossless and lossy compression scenarios.

Index Terms— Light-field, compression, wavelet decomposition, variational optimisation, disparity estimation.

1. INTRODUCTION

Light-field imaging has become an emerging technology nowadays. Compared to traditional photography, light-field photography provides a capability to capture both spatial and directional information of the scene. These rich contents acquired from light-field camera such as plenoptic camera [1, 2], multi-camera array or programmatically moving camera bring a great benefit to numerous applications such as computer vision applications, 3D microscopy, medical imaging and 3D television.

Along with the increasing popularity of light-field photography, the effective coding of light-field images has become an imperative challenge. Recent works on light-field image coding mainly focused on High Efficiency Video Coding (HEVC) framework, either proposed new prediction tools [3, 4, 5] or directly applied HEVC to encode light-field data [6, 7]. Liu et al. [6] proposed to reorder sub-aperture images as a pseudo video sequence and then encode it using HEVC. In [7], Perra et al. applied a similar procedure but instead of using sub-aperture images, they proposed to tile raw light-field image into non-overlapping sub-images. Ricardo et al. [3] proposed two prediction tools to HEVC framework that based on the local linear embedding and self-similarity com-

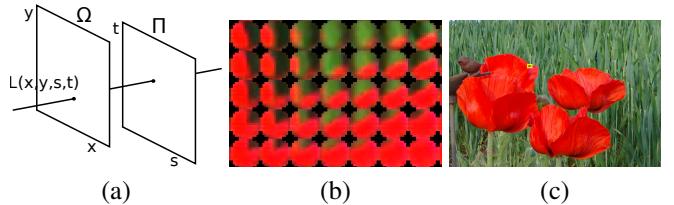


Fig. 1: Lightfield representation and acquisition. (a) Two-plane parameterisation. (b) Lenslet-based sensor image. (c) Central sub-aperture image

pensated prediction. Li et al. [4] worked on an adapted version of bi-directional inter-frame prediction to intra prediction for coding raw light-field image. In [8], Petri et al. proposed a coding scheme that utilises provided disparity map from Lytro software to segment the sub-aperture images and losslessly encode them.

In this paper, we propose a compression framework for light-field images. Our framework decomposes light-field data into multiple views, also known as sub-aperture images, according to the lenslet array structure. A disparity map is then extracted from these sub-aperture images using a variational optimisation approach. Based on this disparity map, an invertible motion-compensated wavelet decomposition scheme is applied to effectively remove redundant information across multiple views. We then employ JPEG2000 to encode all the high-/low-pass sub-band views as well as the disparity map. The framework therefore can support both lossless and lossy compression scenarios.

The remainder of this paper is organised as follows. Section 2 briefly describes a mathematical representation of light-field data and its relation to sub-aperture images. Our proposed compression framework is then discussed in Section 3. We report our experimental results in Section 4 and draw a conclusion in Section 5.

2. LIGHT-FIELD

Light-field could be considered as a 4D parameterisation of the plenoptic function. It is visually described as a ray indexed by its intersection with two parallel planes, as depicted in Figure 1(a).

$$L : \Omega \times \Pi \rightarrow \mathbb{R}, (\mathbf{x}, \theta) \rightarrow L(\mathbf{x}, \theta) \quad (1)$$

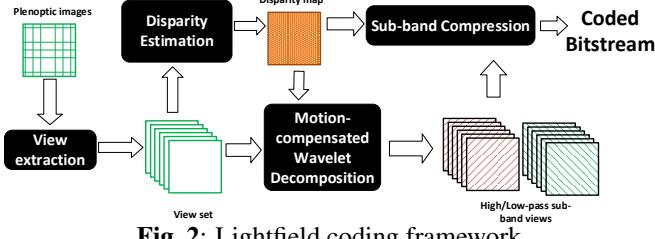


Fig. 2: Lightfield coding framework

where $\mathbf{x} = (x, y)^T$ and $\theta = (s, t)^T$ denote coordinate pairs in the image plane $\Omega \subset \mathbb{R}^2$ and in the lens plane $\Pi \subset \mathbb{R}^2$. For lenslet-based light-field camera, Ω and Π correspond to the spatial and directional coordinates respectively.

Given a light-field L , spatial information could be obtained from one direction by keeping the directional component θ unchanged and varying over all spatial domain Ω . Such spatial information gives us a sub-aperture image (or a view) of the captured scene. The number of sub-aperture images depends on the configuration of lenslet-based camera with a consideration of the trade-off between spatial and angular resolution [2]. Figure 1 (b),(c) show an example of a raw light-field image and a central sub-aperture image extracted from it respectively.

3. PROPOSED CODING FRAMEWORK

Our coding framework is presented in Figure 2. Instead of directly coding raw light-field images, we take advantage of the similarity and redundant information spreading over multiple sub-aperture images. Therefore, the light-field image is firstly decoded into multiple sub-aperture views, which is done by the *View Extraction* module, and a disparity map is then extracted from them by *Disparity Estimation* module. The light-field toolbox developed by Dansereau et al. [9] is used for the first task and a variational optimisation scheme is proposed for the second. The estimated disparity map and all sub-aperture images are sent to the *Motion-compensated Wavelet Decomposition* (MWD) module to extract high-pass and low-pass sub-band views. These sub-band views and the disparity map are then encoded by JPEG2000 in *Sub-band Compression* module. Following sections will discuss in details our disparity estimation algorithm and invertible motion-compensated wavelet decomposition scheme.

3.1. Light-field Disparity Estimation

We model disparity estimation as a variational problem where disparity map $\omega(\mathbf{x})$ is the 2D function that need to be optimized. The general form of this problem is usually written as follows:

$$\operatorname{argmin}_{\omega} E(\omega) = \int_{\Omega} \mathbf{D}(L, \omega) + \alpha \mathbf{S}(\omega) d\mathbf{x}, \quad (2)$$

where \mathbf{D} refers to a data term, \mathbf{S} refers to a smoothness term or a regularisation term and $\alpha > 0$ denotes a smoothness weight. The data function \mathbf{D} penalizes the deviations of

estimated solution ω from the true solution with respect to the constraints on the data L . Such constraints are known as the constancy assumption in optical flow literature [10],[11]. The regularisation function \mathbf{S} takes into account the neighborhood information and guarantees the smoothness of the solution.

Here, the disparity map $\omega : (x, y) \rightarrow \mathbb{R}$ denotes the shift between the pixels in the central view and in its direct east neighbor view. Based on ω we define the constancy assumption across the light-field $L(\mathbf{x}, \theta)$.

$$L(\mathbf{x}, \theta_0) = L(\mathbf{x} + \kappa \theta_i \omega, \theta_i) \quad (3)$$

Here, θ_0 denotes the central view index and θ_i denotes an arbitrary view index within the directional domain Π . $\kappa = \kappa_{\Omega} \kappa_{\Pi} = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$, $k \in \mathbb{R}$, compensates the relative difference between the horizontal and vertical components of directional (κ_{Π}) and spatial (κ_{Ω}) coordinates. Without loss of generality and to simplify the exposition, we will absorb κ into θ .

Data term The Equation 3 actually presents the constancy assumption of pixel intensity that assumes the intensity of corresponding pixels between two sub-aperture images is the same. Suppose ω is small, the first-order Taylor expansion gives:

$$0 = \omega \theta_i^T \nabla_2 L + |\theta_i| L_{\theta} = w^T \nabla_{\theta_i} L, \quad (4)$$

where $\nabla_2 = (\sigma_x, \sigma_y)^T$ denotes the spatial gradient operator, L_{θ} denotes the directional derivative with direction θ_i , $w = (\omega, 1)^T$ and $\nabla_{\theta_i} L = (\theta_i^T \nabla_2 L, |\theta_i| L_{\theta})^T$.

Using Equation 4, we could define a data term that takes into account the intensity constancy assumption:

$$\begin{aligned} D_g(L, \omega) &= \sum_{c=1}^3 \Psi_g \left(\sum_{\theta_i \in \Pi} (L(\mathbf{x} + \theta_i \omega, \theta_i) - L(\mathbf{x}, \theta_0))^2 \right) \\ &\approx \sum_{c=1}^3 \Psi_g \left(\sum_{\theta_i \in \Pi} w^T J_{g,i}^c w \right) = \sum_{c=1}^3 \Psi_g (w^T J_g^c w) \end{aligned} \quad (5)$$

where $J_{g,i}^c = \nabla_{\theta_i} L^c \nabla_{\theta_i}^T L^c$ denotes a light-field motion tensor for a single color channel, and L^c denotes one color channel of the captured light-field L . With it we could define a joint light-field motion tensor J_g^c for all views $\theta_i \in \Pi$ as in Equation. 5. $\Psi_g(s)$ is a positive defined robustification function that helps in reducing the outliers. Here we choose $\Psi_g(s) = \sqrt{s + \epsilon_g}$, with $\epsilon_g > 0$ serves as a small regularisation parameter, it also allows the derivative of Ψ_g available when $s = 0$. This sub-quadratic function was known to be better handling of outliers caused by noise and occlusions [11].

Smoothness term The light-field motion tensor possesses a strong energy in dense texture area, however it is weak in homogeneous area and could lead to wrong estimations. The smoothness term in this case plays a crucial role when it spreads the information through the neighborhood and provides a smooth disparity field. Here, we choose the sub-quadratic penaliser that allows piece-wise smoothness in dis-

parity field.

$$\mathbf{S}(\omega) = \Psi_s(|\nabla_2 \omega|^2) \quad (6)$$

With $\Psi_s(s) = \sqrt{s + \epsilon_s}$, $\epsilon_s > 0$. Combine both data and smoothness terms, we have a final global variational energy function.

$$E(\omega) = \int_{\Omega} \sum_{c=1}^3 \Psi_g(w^T \mathbf{J}_g^c w) + \alpha \Psi_s(|\nabla_2 \omega|^2) d\mathbf{x} \quad (7)$$

This energy function could be optimised using Euler-Lagrange equation. The sub-quadratic functions Ψ_* introduced here will result in a non-linear system of equation. In order to solve it as a linear system of equation, we apply lagged non-linearity method [12] and use Gauss-Seidel iterative solver to estimate the final solution. Detail implementation can be found in our recent work [13].

3.2. Motion-compensated Wavelet decomposition

In our framework, the 2D wavelet transform and the motion compensation are performed jointly in order to effectively exploit the inter-view redundancy of light-field image. The decomposition scheme is based on the lifting implementation of Discrete Wavelet Transform (DWT) [14]. In order to flexibly handle both lossless and lossy compression scenarios, we deploy a two-step lifting scheme based on Haar and bi-orthogonal 5/3 wavelet kernel.

Consider a light-field L as a 2D grid of sub-aperture images defined on directional domain Π . Let $P \subset \Pi$ is a set of view indexes belonging to a row or a column of this grid. We apply the lifting scheme for each view $V_k(\mathbf{x}) = L(\mathbf{x}, \theta_k)$, $\theta_k \in P$. Here $k = 0, 1, 2, \dots$ is indexed from left to right or from top to bottom. A general form of a two-step lifting scheme is formulated in Equation 8.

$$\begin{aligned} h_k(\mathbf{x}) &= V_{2k+1}(\mathbf{x}) + \alpha_1 V_{2k}(\mathbf{x}) + \alpha_2 V_{2k+2}(\mathbf{x}) \\ l_k(\mathbf{x}) &= V_{2k}(\mathbf{x}) + \beta_1 h_k(\mathbf{x}) + \beta_2 h_{k-1}(\mathbf{x}) \end{aligned} \quad (8)$$

Here $h_k(\mathbf{x})$ and $l_k(\mathbf{x})$ are high-pass and low-pass sub-band view, respectively. The weight parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ vary depends on wavelet kernel. For Haar and 5/3 transform, they are set to $(-1, 0, \frac{1}{2}, 0)$ and $(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ respectively.

Let $\mathcal{W}_{i,j}[f]$ denotes the motion-compensated mapping of 2D function $f(\mathbf{x})$ from the coordinate of view V_i to the coordinate of view V_j . The disparity-compensated version of this two-step lifting scheme is as follow.

$$\begin{aligned} h_k(\mathbf{x}) &= V_{2k+1}(\mathbf{x}) + \alpha_1 \mathcal{W}_{2k,2k+1}[V_{2k}](\mathbf{x}) \\ &\quad + \alpha_2 \mathcal{W}_{2k+2,2k+1}[V_{2k+2}](\mathbf{x}) \\ l_k(\mathbf{x}) &= V_{2k}(\mathbf{x}) + \beta_1 \mathcal{W}_{2k+1,2k}[h_k](\mathbf{x}) \\ &\quad + \beta_2 \mathcal{W}_{2k-1,2k}[h_{k-1}](\mathbf{x}) \end{aligned} \quad (9)$$

The inverse transform could be simply derived from Equa-

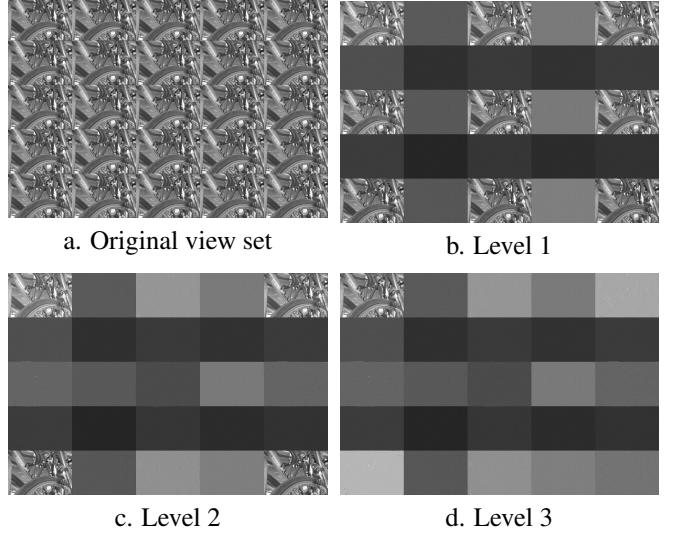


Fig. 3: Demonstration of 3-level 2D MWD for a 5×5 view set of “bikes” lenslet data [15].

tion 9.

$$\begin{aligned} V_{2k+1}(\mathbf{x}) &= h_k(\mathbf{x}) - \alpha_1 \mathcal{W}_{2k,2k+1}[V_{2k}](\mathbf{x}) \\ &\quad - \alpha_2 \mathcal{W}_{2k+2,2k+1}[V_{2k+2}](\mathbf{x}) \\ V_{2k}(\mathbf{x}) &= l_k(\mathbf{x}) - \beta_1 \mathcal{W}_{2k+1,2k}[h_k](\mathbf{x}) \\ &\quad - \beta_2 \mathcal{W}_{2k-1,2k}[h_{k-1}](\mathbf{x}) \end{aligned} \quad (10)$$

For the 2D transform, the 1D transform is first applying horizontally and then vertically on the 2D view set. Figure 3 illustrates three-level MWD of a light-field image with 5×5 directional resolution.

4. EXPERIMENTAL RESULTS

In this section, we compare the compression performance of our proposed coding framework to JPEG-LS and JPEG2000 codecs. Both synthetic and real-world light-field images are considered in our experiment. Five synthetic data are taken from the synthetic 4D light-field dataset [16] and the other five real-world images are taken from JPEG pleno database [15]. The synthetic light-field images have the directional resolution of 9×9 and the spatial resolution of 768×768 (except *horses* with 1152×512 resolution). All the real-world light-field data are captured using Lytro Illum camera that provides 15×15 sub-aperture images and each with 434×625 spatial resolution. Because of the vignetting effect and the geometrical distortion of microlens, the border views are either dark or contain less intensity similarity with the other views. Therefore MWD is applied only for the middle 13×13 views. The remaining views are directly encoded by *Sub-band Compression* module.

Our disparity estimation algorithm is implemented in MATLAB. We applied the Gauss-Seidel iteration solver with successive over-relaxation levels set to 1.88, this allows the solution to converge faster (~ 100 iterations). Figure 4 demon-

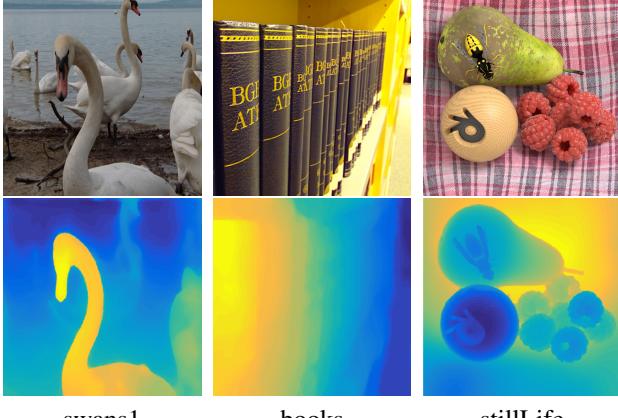


Fig. 4: Estimated disparity map from light-field data. *top* center view images. *bottom* color coded disparity map. Two images from the left are real-world data. The other is synthetic data.

strates the estimated disparity map of 3 light-field images used in our experiment. As could be seen, the disparity maps are smooth but still reserve sharp edges at depth discontinuing area. It is because of the intrinsic sub-pixel precision provided by the continuous formation and the sub-quadratic regularisation.

The lossless compression results are listed in Table 1. The last two columns show the results of our coding framework using Haar and bi-orthogonal $5/3$ wavelet kernel. The other columns report compressed file size of JPEG2000, JPEG-LS and the original light-field data size. From the table, it could be seen that the compressed file sizes produced by our algorithm are smaller than those compressed by JPEG-LS and JPEG2000. In average, it is about 25% and 5.5% better for synthetic data and 11% and 7.3 % better for real-world data compared to JPEG-LS and JPEG2000 respectively. This is attributed to MWD that effectively removes redundant information across multiple sub-aperture images and results in high-pass sub-band views with low energy. From Table 1, it also appears that the $5/3$ wavelet kernel outperforms the Haar kernel for all of test images.

For lossy compression, we set JPEG2000 encoder in our *Sub-band Compression* module to lossy mode. All the sub-band views are encoded with the same compression ratio. The disparity map is however always compressed losslessly. The compression ratios set for this test are 20,40,100,200. Figure 5 compares the average PSNR calculated on YUV color space. We use the equation suggested from JPEG-Pleno call for proposal [17],[18] for this PSNR computation. From the figure, it is apparent that our coding framework with $5/3$ wavelet kernel consistently provides a better compression quality when compared to JPEG2000 in both synthetic and real-world dataset. The gain compared to JPEG2000 increases at lower compression ratio and is even better for luminance component. It could also be seen from the figure that the compression performance when using Haar kernel is not

Table 1: Lossless compressed filesize in mega bytes (MB)

	Image	Raw	JpegLS	Jpeg2000	Haar	$5/3$
synthetic	buddha2	143.33	90.38	62.42	58.04	57.18
	mona	143.33	77.44	52.16	40.99	40.82
	papillon	143.33	70.17	51.57	44.47	44
	stillLife	143.33	113.86	79.59	67.79	66.5
	horses	143.33	96.41	60.8	59.93	58.65
real-world	books	183.09	93.84	103.74	89.18	86.19
	bikes	183.09	106.73	92.14	84.72	78.73
	danger	183.09	105.42	90.41	80.83	74.68
	pillars	183.09	91.48	76.87	71.21	66.74
	swans1	183.09	95.1	88.45	80.39	75.02

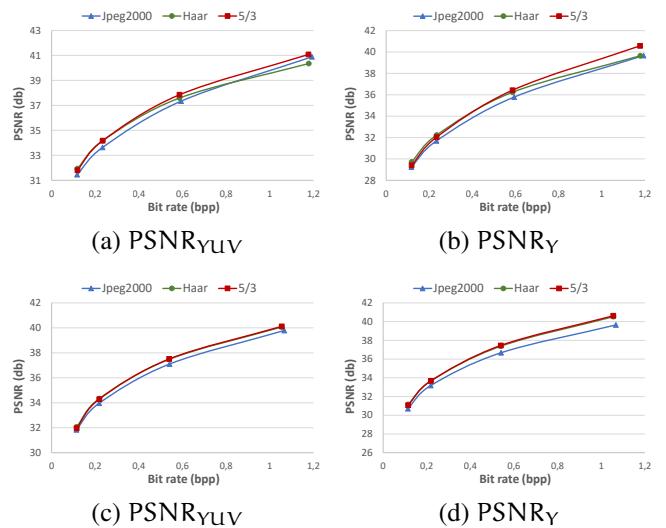


Fig. 5: Average PSNR value for different compression ratio on tested light-field images *top* results for synthetic data. *bottom* results for real-world data.

consistent. It possesses the same compression quality as with $5/3$ kernel on real-world dataset but worse quality in case of synthetic data.

5. CONCLUSION

In this paper, a compression framework based on motion-compensated wavelet decomposition for light-field images has been proposed. We apply a variational optimisation approach to estimate a disparity map from light-field data and use it to warp sub-aperture and sub-band images in a two-step wavelet lifting scheme. Experimental results show that our proposed framework preform better than JPEG-LS and JPEG2000. In the case of lossless compression, our approach provide 5.5% and 7.3% better compression ratio compared to JPEG2000 and 25% and 11% better compression ratio compared to JPEG-LS for synthetic and real-world data respectively. In the case of lossy compression, the result of using $5/3$ wavelet kernel shows consistently better PSNR values compared to JPEG2000.

6. REFERENCES

- [1] Ren Ng, Marc Levoy, Gene Duval, Mark Horowitz, and Pat Hanrahan, “Light Field Photography with a Hand-held Plenoptic Camera,” *Computer Science Technical Report*, vol. 2, no. 11, pp. 1–11, 2005.
- [2] Andrew Lumsdaine and Todor Georgiev, “The focused plenoptic camera,” in *2009 IEEE International Conference on Computational Photography (ICCP)*. apr 2009, pp. 1–8, IEEE.
- [3] Ricardo Monteiro, Luis Lucas, Caroline Conti, Paulo Nunes, Nuno Rodrigues, Sergio Faria, Carla Pagliari, Eduardo da Silva, and Luis Soares, “Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction,” in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. jul 2016, pp. 1–4, IEEE.
- [4] Yun Li, Roger Olsson, and Marten Sjostrom, “Compression of unfocused plenoptic images using a displacement intra prediction,” in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. jul 2016, pp. 1–4, IEEE.
- [5] Caroline Conti, Paulo Nunes, and Luis Ducla Soares, “HEVC-based light field image coding with bi-predicted self-similarity compensation,” in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. jul 2016, pp. 1–4, IEEE.
- [6] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng, “Pseudo-sequence-based light field image compression,” in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. jul 2016, pp. 1–4, IEEE.
- [7] C. Perra and P. Assuncao, “High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement,” in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. jul 2016, pp. 1–4, IEEE.
- [8] Petri Helin, Pekka Astola, Bhaskar Rao, and Ioan Tabus, “Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression,” in *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. jul 2016, pp. 1–4, IEEE.
- [9] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034.
- [10] Andrés Bruhn, Joachim Weickert, Timo Kohlberger, and Christoph Schnörr, “A multigrid platform for real-time motion computation with discontinuity-preserving variational methods,” *International Journal of Computer Vision*, vol. 70, no. 3, pp. 257–277, 2006.
- [11] Henning Zimmer, Andrés Bruhn, and Joachim Weickert, “Optic flow in harmony,” *International Journal of Computer Vision*, vol. 93, no. 3, pp. 368–388, 2011.
- [12] Emanuele Galligani, “A nonlinearity lagging for the solution of nonlinear steady state reaction diffusion problems,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 3, pp. 567–583, 2013.
- [13] Trung-Hieu Tran, Zhe Wang, and Sven Simon, “Variational Disparity Estimation Framework for Plenoptic Images,” in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. Jul 2017, pp. 1–6, IEEE.
- [14] Wim Sweldens, “The lifting scheme: A construction of second generation wavelets,” *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.
- [15] “JPEG Pleno Database: EPFL Light-field data set,” <http://jpeg.org/plenodb/lf/epfl/>.
- [16] Sven Wanner, Stephan Meister, and Bastian Goldluecke, “Datasets and Benchmarks for Densely Sampled 4D Light Fields,” *Vision, Modeling & Visualization*, pp. 225–226, 2013.
- [17] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, “Jpeg pleno: Toward an efficient representation of visual reality,” *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, Oct 2016.
- [18] “Workplan and Specs of JPEG Pleno,” <http://jpeg.org/jpegpleno/workplan.html>.