

© 20XX 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.”

Cserkaszky, Aron, Barsi, Attila, Kara, Peter A. and Martini, Maria G. (2017) Towards display-independent light-field formats. Proceedings of the 2017 International Conference on 3D Immersion (IC3D). Piscataway, U.S. : Institute of Electrical and Electronics Engineers, Inc. ISSN (online) 2379-1780 ISBN (electronic) 9781538646557.

<https://doi.org/10.1109/IC3D.2017.8251903>

TOWARDS DISPLAY-INDEPENDENT LIGHT-FIELD FORMATS

Aron Cserkaszkzy, Attila Barsi

Holografika
Budapest, Hungary
{a.cserkaszkzy, a.barsi}@holografika.com

Peter A. Kara, Maria G. Martini

WMN Research Group
Kingston University, London, UK
{p.kara, m.martini}@kingston.ac.uk

ABSTRACT

The absence of comprehensive and display-independent light-field formats hinders the interchange of light-field content between the capture and display sides of systems, and limits the prevalence as well as the market potential of light-field displays. In this paper, in order to address this issue, we review the shortcomings of the few existing and published formats, and present our considerations regarding expected typical use cases and research directions towards the goal of display-independent light-file formats. As a major contribution of this paper, we propose a novel display-independent light-field file format.

Index Terms— Light-field file format, light-field display, display independence, light-field conversion.

1. INTRODUCTION

File formats specifically designed for storing light-fields are most certainly not new in the field. However, the relative prevalence of light-field capture systems and virtual light-field rendering technologies – compared to real light-field displays – created an imbalance in the focus of current formats. Existing light-field formats are either designed to represent dense and uniformly sampled light-fields or they contain a set of images captured from different positions by perspective cameras along with calibration parameters. At the time of this paper, it can be stated that only little work has been done towards display-focused light-field formats that enable the efficient interchange of light-field content between capture and display systems.

In order to display light-fields in the existing light-field formats on a light-field display, a computationally expensive conversion needs to happen. This conversion operation computes the appropriate color value for all the display rays that can be emitted from the light-field display, using the captured light-field rays in the input data. This also implies a noticeable loss of quality when the sampling of the input data is not dense enough to provide a measured color value from nearby capture light-field rays for all display light-field rays. Such loss of visual quality can result in the severe degradation of the Quality of Experience (QoE). Another implication

of the conversion is that the necessary computation time likely makes real-time streaming infeasible.

It is important to note here that while *all* displays are effectively light-field displays – albeit with very limited directional views – the term is generally used when the angular resolution of the display is high and the viewer experiences continuous motion parallax. Such light-field displays already exist on the market [1][2] and their characteristics include a highly nonuniform – or somewhat angularly uniform – display light-field, that is rarely taken into account by existing work in this field.

In this paper, we introduce our proposal on a display-independent light-field format. The goal of the proposed format is to be able to represent all possible light-fields through the usage of a middle layer. The paper highlights the potential gains and advantages of the successful implementation and utilization of such a format.

The remainder of this paper is structured as follows: Section 2 provides an overview of the state-of-the-art light-field systems, utilized techniques and use cases. Section 3 briefly discusses the light field formats available at the time of this paper. We propose our own novel format in Section 4, followed by the further relevant research directions in the field in Section 5. The paper is concluded in Section 6.

2. LIGHT-FIELDS, DISPLAYS AND CONVERSION

2.1. General remarks

Light-field [3], in the broadest sense, is a seven-dimensional function that describes the intensity of light rays starting from every point in space within a given domain, going to every direction within a field of view (FOV), on all described frequencies, and its change over time. Measuring, rendering or viewing this entire light-field is not realistic, therefore the plenoptic function [4] was described to reduce the required amount of dimensions by using some assumptions about light and the Human Visual System (HVS). The first assumption is that light rays normally traverse air without any change of intensity, thus making one of the spatial dimensions redundant. In other words, instead of describing a light-field scene with a volume, we can simply consider rays on a convex sur-

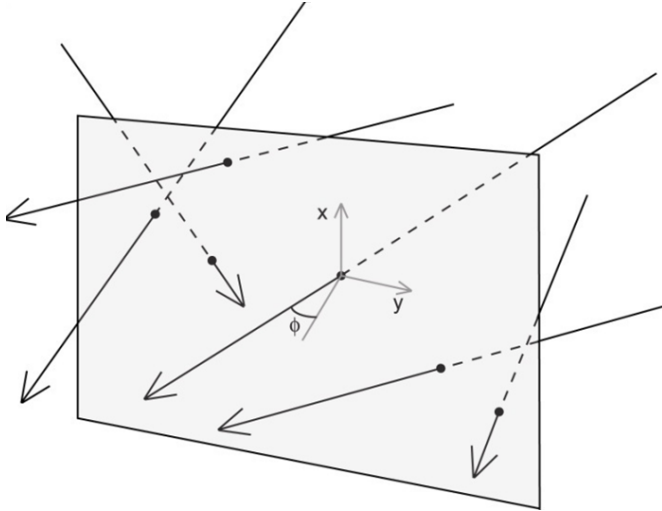


Fig. 1: Illustration of light-field rays creating the “window of light”.

face around it. The second assumption is that our human eyes perceive light on the RGB color space, thereby frequency as an independent dimension is replaced by three separate – yet structurally identical – light-fields of the color red, green and blue. This leaves us with four non-temporal dimensions, that fundamentally describe the structure of light rays in the light-field.

This 4D plenoptic function in a simple sense is akin to the concept of the “window of light”, illustrated on Figure 1, where we may change our position of observation both horizontally and vertically, and we would see the perspectively correct 2D images coming from the plane of the window, with each of our eyes. As it is similar to looking through an actual window in reality, our brain is able to perceive the scene behind the plane in true 3D. An integral part of the light-field visual experience is the smooth continuous motion parallax, which means that the observer moving in front of the window sees the smoothly changing views of the scene, in which the viewing of objects closer to the window changes faster.

Another optional reduction of this light-field is the restriction of horizontal movements of the observer. This is actually practical since humans tend to move horizontally, and more often look at objects or scenes from many sides than from above and below. This reduces the plenoptic function by another dimension and creates the horizontal parallax only (HPO) light-field. The unrestricted case – when both horizontal and vertical directional views are available – is called the full parallax light-field. In both these cases though, we refer to the allowed horizontal movement within the field of view as the baseline. It is important to note here, that motion towards or against this window also induces small changes in the vertical dimension, through the change of perspective. Figure 2 shows such a HPO display and how it creates the 3D visual experience.

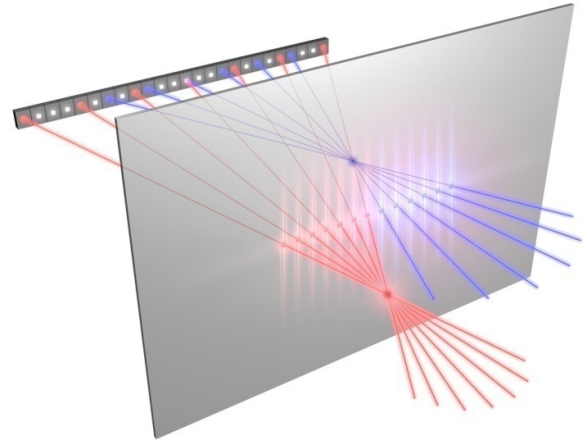


Fig. 2: The principle mechanism of a light-field display. Observers perceive the red and the blue points at the junction of the respectively colored rays.

The primary function of displays is to generate light rays that – after those rays are sampled by the HVS and then processed by the brain – create meaningful information to the user. In a general sense, most of our display technologies can be considered light-field displays, as they all are able to recreate portions of the plenoptic function. However, our definition of light-field displays is based on three essential requirements. The first is that the display needs to change the displayed light-field with a frequency consistent with the perceptual capabilities of the HVS, so at least 24 frames per second. The second is that the displayed light-field has to have wide baseline, meaning more than 1 meter, in order to allow the user to move in front of the display, or a group of users to experience the same scene from multiple angles. The final – and most important one – is the requirement of smooth motion parallax, that in practice translates to angular views in the range of hundreds in each direction (horizontal and vertical). This clearly defines the boundary between light-field displays and multiview displays. The latter ones typically have 2 to 10 horizontal views to display and the user perceives jumps between these. Current high-end light-field displays on the market emit light rays in the order of hundred millions to create the many views necessary for smooth motion parallax, and in some sense light-field displays can be considered as “manyview” displays.

Moving to the capture side, the creation of light-fields can happen from virtual or real sources. A virtual source is for example an animated 3D scene or a game, where views are rendered with virtual cameras or an arbitrary amount of individual rays are ray-traced. This is a fairly straightforward process, but requires considerable computational resources to perform due to the high overall resolution. Real light-field capture sources are typically arrays of regular digital cam-

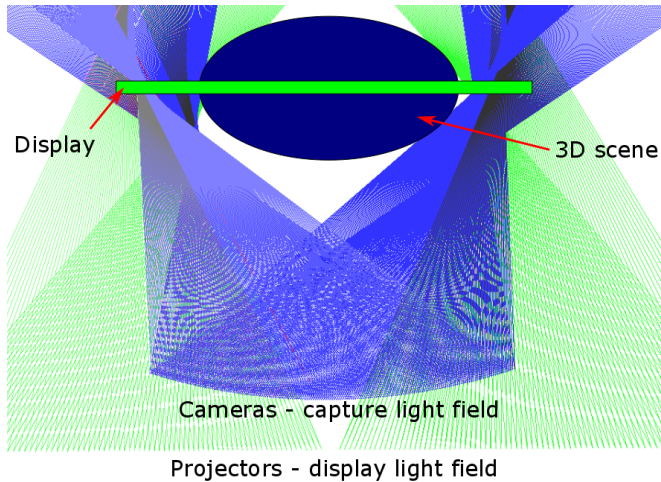


Fig. 3: Overlay of a capture light-field (blue) on a display light-field (green), illustrating the challenge of conversion. Note the inefficient use of the display’s FOV.

eras. This poses challenges on many fronts. First of all, due to the physical size of the cameras, they often cannot be sufficiently tightly stacked next to each other to provide a smooth motion parallax experience. The cost of a dense camera rig is also close to prohibitive and makes them impractical in many use cases. This often necessitates the interpolation of virtual cameras [5] or the reconstruction of the full light-field [6]. Another challenge is that we need the precise geometrical properties of each ray supplied by the cameras. To achieve this, we need to calibrate the camera rig, where each camera’s position, orientation, lens distortion, color balance and other properties are estimated. This is already a very difficult task on its own, but we also have to ensure that the properties of the camera rig stay constant with time, taking into consideration external temperature variations and minor mechanical incidents.

Conversion is the process that restructures a given light-field into a new one with the desired ray structure. Figure 3 shows such a scenario with the two distinct light-field structures are overlaid on each other. Often we have a light-field captured by camera arrays, that have many hundreds of millions of combined pixels, but the light rays corresponding to these do not directly match with the rays we are able to emit from the display, therefore we must use the operation of conversion. What conversion practically does is that it searches through the capture light-field to find rays that can be used to interpolate the the rays of the display light-field. This is a computationally difficult task due to the 4D search space and one-to-many relations between the light-field rays, and the quality of the converted light-field depends greatly on the level of match between the sampling of the ray space of the two light-fields.

2.2. Use Cases of Light-Field Displays

2.2.1. Requirements

The true glasses-free 3D visual experience that light-field displays can provide substantial benefits to countless applications and creates new ones, but in the context of light-field formats, let us concentrate on their technical requirements. In general, we have to consider three important technical factors.

The first one is the available bandwidth, or in general the storage space required for a given light-field. This could be analogous to the challenges tackled by traditional 2D streaming and video encoding technologies, however, the added angular domain of light-field inflates the required amount of data by two magnitudes, even for HPO displays, and this makes existing solutions inefficient. The second technical factor is the delay, which is the time required from capturing the light-field to being able to perceive it on the display. This is also similar to existing streaming applications, but the novel issue here is the potentially large amount of computational resources required by the conversion process, that creates the displayable light-field from the captured one. We will discuss the process of conversion in detail later in this paper. The final factor is the issue of the computational resources themselves, that are closely intertwined with both the bandwidth and the delay. Reducing the necessary bandwidth requires complicated light-field compression operations on the capture side and decompression on the display side. Furthermore, maintaining a low delay needs the compression and other operations like conversion to happen at near instant speeds.

It is important to note here, that in this paper we do not consider quality as a separate factor, but assume that the data volume is such that it sufficiently satisfies the quality criteria posed by the specific use case.

2.2.2. Use cases

Considering these criteria let us categorize the use cases of light-field displays based on whether the capture of display side of the processing happens live or off-line.

Light-field cinema is a novel field, and has the unique advantage that it can play a light-field video recreating the experience similar to a theatrical presentation, where all viewers see a different angle of the scene based on their positions. A high-resolution light-field cinema system could approach having a billion pixels, but neither the capture, nor the display side processing is limited by time constrains. The captured light-field could be post-processed to have the highest quality without regards to computational resources or storage (bandwidth) requirements. The conversion of these light-fields to each utilized light-field cinema display model is also not an issue, since these displays will remain limited in type.

Light-field teleconferencing is the almost exact opposite in terms of requirements. Low delay and handling limited bandwidth are the key factors for teleconferencing. This means

that despite the available computational resources, the captured light-field cannot be post-processed and must be converted quickly to the display light-field, as well as compressed with the fastest codec and transmitted through the network. The main issue here – besides the obvious computational challenges – is that the capture side will only be able to quickly either convert to one specific display type, or send the compressed camera images, which have to be converted at each display participating in the teleconference. The latter is a wasteful and inefficient model, and the former promotes incompatible and closed systems. General live-streaming is very similar to the use case of teleconferencing, albeit with relaxed delay requirements and with an even greater need for a display-independent light-field format.

The use case of light-field video-on-demand (VoD) – or light-field-on-demand (LFoD) – service provides a mixture of the two formerly mentioned applications. Light-field can be captured, post-processed, converted and compressed in an arbitrary amount of time, however, the resulting light-field must be optimized for size and quality for all display models on the market in order to allow these to instantly stream the content and display it, preferably with small delay and without computationally expensive steps. The option of converting the content to all possible light-field displays is extremely wasteful regarding storage space due to the large data sizes, and also disallows peer-to-peer delivery of the stream between different display models.

The fourth application category is the capture-to-cloud use case, where a lightweight capture-only device – that has no (or not large enough) storage space or significant computing power – captures light-field and streams it continuously to a cloud storage or dedicated light-field processing service. In this case fast conversion and post-processing is not required and the use case is less relevant for display independent formats.

There are many more use cases of light-field displays than these, but the live-streaming and LFoD application categories are unique in the sense that display-independent light-field formats greatly inhibit their potential success. We will focus on these while reviewing existing formats and then propose a display-independent light-field format.

3. EXISTING LIGHT-FIELD FORMATS

In this section we discuss existing light-field formats with focus on the structural match between their described light-field, and light-fields that we can create with current displays. It is important to note here that it is possible to build a light-field display specifically to represent a certain light-field format's ray distribution, but this would impose technological restrictions on light-field displays and likely increase their cost by a large margin. This is infeasible in such a nascent field and would unduly reduce their market potential.

3.1. The *.lif* format

The work of Levoy and Hanrahan – already in 1996 – extended into specifying a light format based on the very general concept of slabs of bi-planar (s,t,u,v) coordinate-based ray parametrization [7]. Their format essentially contains the geometric description of two planes that the described rays intersect, and these intersection points are described with the 2D s,t coordinate pair on the first plane, and u,v on the second [8]. A single one of this bi-planar construct (slab) is not able to describe all possible rays, however, multiple slabs can. This sampling is uniform in all four of its coordinates, which is optimal only in the scenario of a likewise spatially uniform display light-field sampling. No such light-field display exists on the market at the time of this paper.

3.2. The Lytro format (*.lfp*)

The Lytro cameras are the only commercial light-field products that had actually successfully entered the consumer market. Both marketed versions capture narrow baseline light-field images with their microlense-based optics. Lytro also released a software package to process and edit the captured light-field images. The raw file produced by the camera contains the microlense images. This can be processed with the provided software into an RGBD (color+depth) total focus image [9]. The inherently narrow baseline of the RGBD image format means that it is not a suitable format for the typically wide-baseline light-field displays.

3.3. General camera image set formats

Light-field is often captured with a camera system, usually consisting of consumer-grade perspective cameras, where either many cameras capture images from the scene at the same time, or a few cameras take images of a static scene from multiple known positions and directions. An obvious general light-field format for this scenario is to just store the camera images with their calibration parameters, including the intrinsic, extrinsic and distortion parameters of the cameras [10] [11] [12]. The existing sophisticated image compression algorithms make this format very efficient in the aspect of storage space. However, the limited number of camera positions – which are usually distributed along one line – make the sampling distribution of these capture light-fields very dissimilar to the sampling distribution of the display light-fields that we aim to represent. This would necessitate not only the conversion but also light-field interpolation or reconstruction to be performed on such datasets.

3.4. JPEG Pleno

The JPEG collaboration is currently in the process of developing a standardized light-field format within the frame of

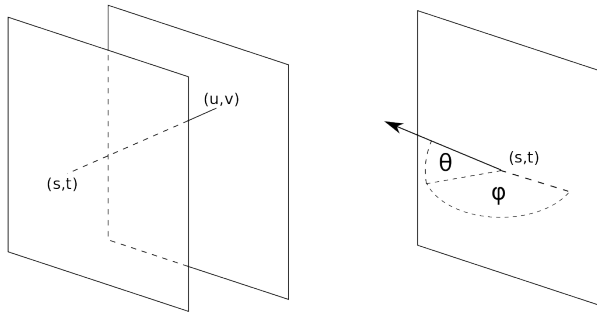


Fig. 4: The bi-planar (s,t,u,v) parametrization and the proposed angularly continuous (s,t,φ,θ) parametrization.

the JPEG Pleno framework [13]. The efforts target both high-density camera arrays and lenslet-based camera systems. This work is very promising, but the focus on the compressed storage of short-baseline light-fields captured by such lenslet-based cameras has certain limitations. Attempting to display these short-baseline light-fields on wide-baseline light-field displays is a very inefficient use of the displays' FOV, and the current scheme cannot efficiently scale up to camera-array-based capturing setups.

4. PROPOSED DISPLAY-INDEPENDENT FORMAT

A display-independent light-field format is a somewhat elusive goal, as it aims to contain the native display light-field ray distribution for multiple – potentially very different – displays. So a truly display-independent format, that is useful in many use cases, needs to be a flexible and general middle, or a “mezzanine” format, that facilitates the efficient interchange between capture and display light-field, through a middle layer. This is analogous to the use of the equirectangular 2D representation of 360-degree video applications, where the efficient compression is solved with existing technologies and the 2D representation is independent from the specific resolution of the virtual reality display.

We have already discussed the shortcomings of the bi-planar light slab parametrization. Yet the mezzanine layer, between capture and display light-field, needs to be similarly general to represent all possible light-fields. The plane on one side is likely optimal, as the screen of the light-field displays tend to be flat and the crossing points of their light-field rays are close to being uniformly distributed, but on the other side, their light-field distribution is more compactly quantized with a section of a cylinder, or in the case of full parallax displays, a section of a sphere. This creates an angularly continuous sampling of rays originating from the equidistantly sampled plane, illustrated on Figure 4, next to the bi-planar parametrization. The proposed format's two spatial and one

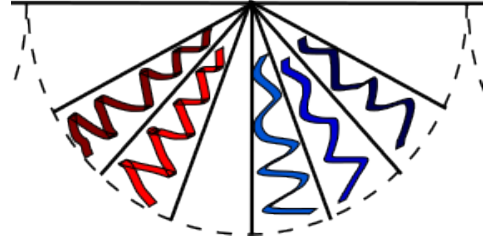


Fig. 5: The proposed angularly continuous representation showing the color information in angular bins.

or two angular parameters of the light-field can be independently set, based on quality requirements or available resources. This has the added benefit that – just like nowadays with normal images – we have realistic quality expectations when we view an image with a certain resolution, on a display with another resolution. This would also help display manufacturers to market their products as this general light-field parametrization, along with the FOV, provides objective parameters for the general comparison of different displays.

This general format – while describing a light-field matching to no real display – allows such use cases when the data is streamed to multiple different or unknown displays at the same time, and the proper conversion of the captured light-field to the display light-field is not possible or feasible. Another benefiting use case is when the availability of computational resources is low on the capturing side, either because of time constraints due to real-time streaming, or by limitations of portable capturing systems regarding processing power. The compressibility of a sparse general representation would also be useful when the captured light-field contains overlaps or other inefficient sampling effects, so a low-bandwidth use case would also benefit from such a middle format. On the display side of the transmission, the mezzanine format is similarly useful in a streaming scenario. A display-specific conversion look-up table could on-the-fly convert the fixed middle format to the display light-field. This also bypasses the know-how or trade secret issues that would potentially arise, if the display manufacturers have to publish the detailed specifications of their displays light-field sampling distribution in order to gain access to light-field content. Generating this conversion look up table is fairly straightforward so the display will be able to display light-field streams with dynamic changes of resolution.

Another general advantage of such a format is that the inherent structure allows the easy interpolation of missing light rays from the neighboring ones. This interpolation will decrease the quality of the visual experience to some degree, much like in the case of a zoomed in normal 2D image, however, it will at least retain the smooth motion parallax, as the viewer will not see jumps between the views. Figure 5 shows the angular color rays emitted by one spatial pixel of a light-field display.

5. RESEARCH DIRECTIONS

The visualization quality enabled by the proposed display-independent format needs to be evaluated with the usual bandwidth, delay, and computational requirements parameters, utilizing both of the available objective and subjective quality assessment methodologies as a first step. These tests need to provide data points to the typical parameters of the discussed applications. The end goal is to establish the optimal spatial and angular resolutions for the appropriate use case scenarios.

The proposed format should also be tested in the context of Quality of Experience. The quality switching between higher and lower resolutions of the format needs to be as least intrusive as possible to gain adoption with viewers and light-field content providers.

Light-field in such format can be easily compressed and decompressed using the different angular views as separate images (which is a matrix in case of full parallax data), however, even more efficient, joint compression schemes need to be developed before the widespread use of light-field technologies. These compressions also have to be evaluated with perceived quality in mind.

We have already discussed the largely analogous challenges of horizontal parallax only and full parallax light-fields, yet the additional 1-2 orders of magnitudes in data volume on top of the already quite large HPO light-fields might necessitate paradigm shifts in processing, storing and transmitting such datasets.

A related research field in 5G communications [14] (and beyond) is the reliable and QoE-aware transmission of such immense data volumes. Light-field formats probably will not be integral to solving these challenges, but we must mention it as an important research area and an essential part of light-field content streaming in the future.

6. CONCLUSIONS

In this paper, we presented our considerations and contribution to create a display-independent light-field format, that can be used in typical use cases that are expected from such a display technology. We conclude that the existing work in light-field file formats is not sufficient to achieve such goal. We proposed research directions that we expect to be relevant and important towards display-independent formats. Furthermore, we introduced our own novel proposal for a light-field file format. Utilizing a middle layer for the efficient interchange between capture and display light-field satisfies the requirement of being sufficiently general in order to represent all possible light-fields. In our proposal, light-fields are described with two spatial and one or two angular parameters, which comes with multiple benefits. In future works, we aim to publish our researches evaluating the performance of the middle format regarding visualization quality with objective and subjective metrics.

7. ACKNOWLEDGMENT

The research in this paper was done as a part of and was funded from the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 676401, ETN-FPI, and No 643072, Network QoE-Net.

8. REFERENCES

- [1] Holovizio 80WLT light field display. http://www.holografika.com/Documents/HoloVizio_80WLT-emailsize.pdf (retrieved September 2017).
- [2] Real Eyes 3D display. <http://real-eyes.eu/3d-technology/> (retrieved September 2017).
- [3] Andreĭ Gershun. The light field. *Studies in Applied Mathematics*, 18(1-4):51–151, 1939.
- [4] Edward H Adelson and James R Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, Cambridge, MA, 1991.
- [5] Aron Cserkaszy, Attila Barsi, Peter A Kara, and Maria G Martini. To interpolate or not to interpolate: Subjective assessment of interpolation performance on a light field display. In *Multimedia & Expo Workshops (ICMEW), 2017 IEEE*, pages 55–60. IEEE, 2017.
- [6] Suren Vagharshakyan, Robert Bregovic, and Atanas Gotchev. Light field reconstruction using shearlet transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [7] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.
- [8] Stanford light field file format. https://graphics.stanford.edu/software/lightpack/doc/file_format.html (retrieved September 2017).
- [9] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma. Lytro camera technology: theory, algorithms, performance analysis. In *Proc. SPIE*, volume 8667, page 86671J, 2013.
- [10] Disney Research light field dataset. <https://www.disneyresearch.com/project/lightfields/> (retrieved September 2017).

- [11] Nagoya University sequences.
<http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/>
(retrieved September 2017).
- [12] Heidelberg Benchmark 4D light field dataset.
<http://hci-lightfield.iwr.uni-heidelberg.de/>
(retrieved September 2017).
- [13] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, and Peter Schelkens. JPEG Pleno: Toward an efficient representation of visual reality. *IEEE Multimedia*, 23(4):14–20, 2016.
- [14] Jonas Hansen, Daniel E Lucani, Jeppe Krigslund, Muriel Médard, and Frank HP Fitzek. Network coded software defined networking: enabling 5G transmission and storage networks. *IEEE Communications Magazine*, 53(9):100–107, 2015.