Compressive HDR light field imaging using a single multi-ISO sensor

Ehsan Miandji, Hoai-Nam Nguyen, Saghi Hajisharif, Jonas Unger, Christine Guillemot

Compressive HDR light field imaging using a single multi-ISO sensor

Ehsan Miandji†, Hoai-Nam Nguyen‡, Saghi Hajisharif†, Jonas Unger† and Christine Guillemot‡, *Fellow, IEEE*
†: Linkoping University, Norrköping, Sweden
‡: Inria Rennes – Bretagne-Atlantique 263 Av. Général Leclerc, 35042 Rennes Cedex, France

***Abstract*—In this paper, we propose a new design for single sensor compressive HDR light field cameras, combining multi-ISO photography with coded mask acquisition, placed in a compressive sensing framework. The proposed camera model is based on a main lens, a multi-ISO sensor and a coded mask located in the optical path between the main lens and the sensor that projects the coded spatio-angular information of the light field onto the 2D sensor. The model encompasses different acquisition scenarios with different ISO patterns and gains. Moreover, we assume that the sensor has a built-in color filter array (CFA), making our design more suitable for consumer-level cameras. We propose a reconstruction algorithm to jointly perform color demosaicing, light field angular information recovery, HDR reconstruction, and denoising from the multi-ISO measurements formed on the sensor. This is achieved by enabling the sparse representation of HDR light fields using an overcomplete HDR dictionary. We also provide two HDR light field data sets: one synthetic data set created using the Blender rendering software with two baselines, and a real light field data set created from the fusion of multi-exposure low dynamic range (LDR) images captured using a Lytro Illum light field camera. Experimental results show that, with a sampling rate as low as 2.67%, using two shots, our proposed method yields a higher light field reconstruction quality compared to the fusion of multiple LDR light fields captured with different exposures, and with the fusion of multiple LDR light fields captured with different ISO settings.**

***Index Terms*—HDR, light fields, data sets, compressive sensing**

I. Introduction

Efficient, high quality capture of light fields is a key research challenge in the development of new imaging techniques. By capturing both the spatial and angular variations in the light incident onto the sensor and encoding this as a 4D data set, a light field enables detailed capture, representation and analysis of scenes and objects, and is recognized as an underpinning technology in computational photography. Light field imaging has enabled a range of research and emerging applications in areas ranging from computer graphics [1], [2], [3], [4] and everyday photography [5] to computer vision [6], [7], microscopy [8], [9] and displays [10] to name a few. Due to the complexity inherent to light field capture, current cameras and imaging systems are limited to low dynamic range (LDR) capture. High dynamic range (HDR) light field capture still requires multiple exposures to cover the desired dynamic range.

Designing a compact light field camera is a difficult task as there is a trade-off between spatial resolution, angular resolution and visual quality. Current systems are either based on non-portable camera arrays [11], [6], or a single sensor with a lenslet array [12], [5], or a coded mask [13], [14], [15], [16] placed in front of the sensor. Effort has been in parallel dedicated to the design of 2D HDR imaging solutions, based on new sensor technologies [17], [18], by merging multiple images captured at different exposures with a standard camera, or by placing optical elements with spatially varying transmittance in front of the image sensor [19], [20].

The problem of HDR light field acquisition using a 2D sensor remains an open and challenging problem despite the recent advances in both 2D HDR imaging and compressive LDR light field acquisition. The main challenge here is indeed the reconstruction of an HDR light field from a single LDR image recorded on a monochrome sensor equipped with a Color Filter Array (CFA). This single monochrome image hence should encode not only the HDR information of the scene, but also angular and spectral measurements of the light field. To address this problem, we introduce a novel framework for compressive capture of HDR light fields combining multiple ISO photography with mask-based coded projection techniques. To solve the HDR imaging problem, a possible approach would be to place a filter of spatially varying transmittance in front of the sensor in order to record different exposures in adjacent pixels of the sensor, as in [19] for 2D HDR imaging. However, this solution would require increasing the exposure time to compensate for the light attenuation induced by the filter. Instead, we consider a design where a color coded mask is placed in front of a multi-ISO sensor with per-pixel control over the ISO value, e.g. [21], or with solutions such as the "magic lantern" firmware for dual-ISO HDR photography [22]. In order to improve the quality of the reconstructed HDR light field, we can increase the number of recorded measurements by increasing the number of shots and apply the same reconstruction method jointly on all the captured shots.

Our compressive HDR light field imaging framework captures a coded image with a varying per pixel gain encoding the scene. The sensor image captured through the mask, the varying per pixel gain, and the CFA, encodes spatial, angular, and color intensity variations in the scene. This coded projection image compresses the incident scene radiance information such that the full HDR light field can be recovered as a tractable inverse problem. To solve this, we present a novel joint spatio-angular-HDR reconstruction algorithm using a trained dictionary specifically designed for HDR light field reconstruction. The joint reconstruction includes a confidence matrix based on the pixel intensity and acquisition noise, effectively performing denoising as an integral part in the reconstruction.

This work was supported by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

In order to evaluate our approach we have created an HDR light field benchmark data set that will be made publicly available. The data set consists of synthetic and real HDR light fields rendered in Blender or captured with a Lytro Ilum camera using multiple exposures in a variety of challenging lighting conditions.

The main contributions in this paper can be summarized as:

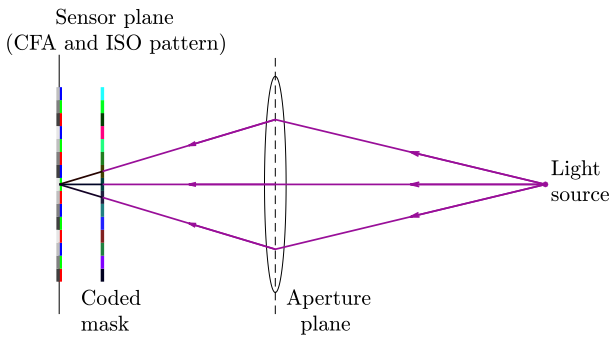
- A novel single sensor camera design for capturing HDR light fields, combining multi-ISO photography and coded mask acquisition principles.
- A novel joint spatio-angular-HDR light field reconstruction algorithm that is robust to noise.
- A novel over-complete dictionary tailored to HDR data.
- A novel HDR light field data set with synthetic and natural image data intended for evaluations and performance benchmarks.

To evaluate our approach we perform an investigation of different per pixel ISO/gain patterns and sensor parameter configurations. More precisely, we consider three ISO patterns, a full random pattern as well as more structured block-based and row-based interlacing patterns. We report results for each pattern in order to investigate whether full randomness, due to increased incoherence with the dictionary, could improve the reconstruction performance. While we observed improved performance for random per-pixel ISO, we acknowledge the fact that these types of sensors have not reached the commercial market yet, and that row-based and block-based patterns enable cheaper and easier hardware implementations; the row-based interlacing ISO pattern has been utilized for HDR imaging using off-the-shelf consumer cameras [23].

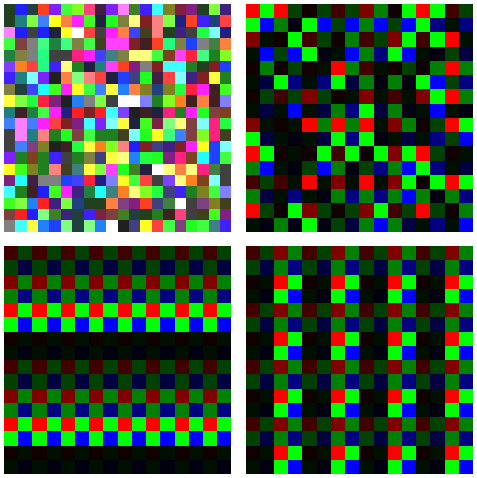
To expand the analysis of our method, we assess how the proposed approach performs when increasing the number of captures, referred to as shots in what follows, and we assess its robustness to noise. The results show that the proposed camera architecture and the joint spatio-angular-HDR yields a higher reconstruction quality as compared to fusing multiple LDR light fields, captured with different exposures, using traditional methods [24]. In other words, our method outperforms traditional methods based on exposure bracketing given only 2.6% of the total number of pixel samples (assuming two shots). We also compare our results with a two step method of recovering the light field from the 2D compressed measurements followed by the reconstruction of the HDR values using a state-of-the-art deep learning method [25] on each reconstructed view.

II. Related Work

a) HDR imaging: In recent years, several techniques and architectures have been proposed for extending the image sensor dynamic range, by using a computational element that measures the pixel saturation time, i.e., the time to attain full potential well capacity, [27]. However, such detection of saturation for each pixel is not easy to implement. Other CMOS sensor designs have been proposed with multi-bucket pixels enabling time-multiplexed exposures [21], by modifying the readout circuits [18], [17] to improve the pixel dynamic range, or by adapting their sensitivity [28]. A quantitative study of HDR image sensors can be found in [29].



(a) Camera architecture



(b) Coded mask and three examples of ISO patterns

Fig. 1: (a) The camera architecture is composed of a main lens, a coded mask, and a photo-sensitive sensor equipped with a CFA on top (the right side of the sensor plane) and a ISO pattern behind (the left side of the sensor plane). (b) Examples of coded masks (top-left) and three ISO patterns: *Random*, *Row-based*, and *Block-based* overlaid on top of the Bayer CFA pattern. These images are included for illustrative purposes and do not represent an attenuation mask since the gain is applied after the acquisition and that the gain/ISO amplifies the signal rather than attenuating it.

Alternative approaches have been proposed for HDR imaging acquisition that consist of placing an optical mask with a spatially varying transmittance in front of the image sensor. The mask allows recording different exposures of the scene on adjacent pixels of the sensor [19]. Spatially varying sensor gains have also been considered in [22] for capturing high dynamic range (HDR) images in a single shot. The HDR image is reconstructed using an algorithm based on the sensor noise model to suppress the image noise and remove saturated values [30], [31], [32]. The authors in [33] and [34] propose an architecture that is independent of the sensor, in which the light is split using some optical element, e.g. a pyramid-shaped mirror or a refracting prism, and redirected toward a set of sensors associated with absorptive filters to produce images with different exposures. In [35], an optical design is proposed with a rank-1 phase pattern, where the diffractive optic point spread function (PSF) is jointly optimized with the reconstruction method that recovers the latent HDR scene from the input measurement. A joint design of a spatially-varying modulation mask and of the HDR image reconstruction network is proposed in [20].

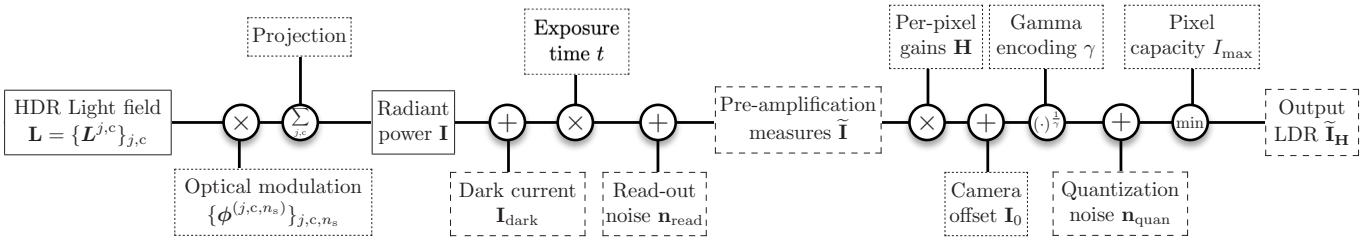


Fig. 2: The imaging pipeline corresponding to the camera architecture shown in Fig. 1 represents the optical projection of light field data on the sensor and the digitization at the sensor using a noise model inspired by [26] to take into account different gain variations.

Another type of approach for HDR imaging, following the principles of epsilon photography [36], involves taking a series of images with different exposure times with a standard camera through exposure bracketing, and then combining them into an HDR image using a hat function [24] or fusion weights based on the statistical characteristic of the sensor noise [26], [37]. The merging can also be performed using either patch-based methods [38], [39], [40], low rank matrix completion (LRMC) [41], [42], [43], [44], and more recently deep learning methods [45]. A model of per pixel intensity variations, observed when the camera parameters are varied, is proposed in [36], inspired by the confocal stereo method [46].

Methods also exist to reconstruct an HDR image from one single LDR image by hallucinating missing information in under and over-exposed regions, using either convolutional neural networks [47], [48], [49], [25], or convolutional sparse coding [50]. A comprehensive review of the subject is provided in [51].

Even though the proposed design is related to some of the above approaches, in particular using spatially varying exposures for capturing HDR images, here we consider multi-ISO sensors together with a coded mask that allows us to compress a light field in both the spatial and angular dimensions.

b) Light field imaging: Over the past decades, a range of acquisition devices have been developed for efficient capturing of 4D light fields, from multi-view imagers [1], [11], [6] to multi-view coders which encode angular information of 4D light fields onto 2D sensor images from which the light field views can be recovered. More compact acquisition devices have been designed by modifying the architecture of conventional cameras, e.g. using micro-lens arrays (MLA) placed on the optical path before the camera sensor [12], [5], to capture the light field passing through the main lens.

Alternatively, coded mask designs have been considered instead of MLA to modulate 4D light fields into 2D projections on the sensor. Different reconstruction algorithms can then be used to restore the light field from its 2D coded projection [13], [14], [15]. The light field acquisition and reconstruction problems are formulated by a compressive sensing framework, in which the sensing matrix is materialized by a coded physical mask. Thanks to the use of a coded mask, instead of recording a spatial multiplex of 2D slices of the light field, as in micro-lens based camera architectures, the photosensor records a set of linear measurements from which a higher resolution light field can be reconstructed. Babacan et al. [52] place a randomly coded mask on the aperture plane to obtain incoherent measurements of the light field. Multiple shots are captured as random linear combinations of angular images by separately opening one region of the aperture and blocking the light in the others. Marwah et al. [14] propose to place a monochrome mask between the aperture and sensor planes to record optically coded projections on a single image sensor, while Miandji et al in [53] and [15] use respectively a random binary mask or a moving color coded mask affixed to the sensor to extract incoherent measurements.

While the previous methods rely on patch-based dictionaries with a sparsity assumption for light field reconstruction from the set of sparse measurements, Heide et al. [54] employ learned convolutional sparse coding features to synthesize light field views. Light fields can also be reconstructed from a set of measurements using convolutional neural networks (CNN) [55], [56], [57], given a pre-defined coded mask. In contrast, Inagaki et al. [58] optimizes the mask pattern together with the parameters of the reconstruction algorithm in an end-to-end auto-encoder learning framework. Moreover, a learned convolutional network architecture is used in [59] to compute the coded sub-aperture images, from which the light field is reconstructed using an iterative optimization approach with a deep spatio-angular regularization prior. Note that an attenuating mask placed in the optical path of a lens-based camera is considered in [60] to compute refocused images at full sensor resolution for layered Lambertian scenes.

c) HDR light field imaging: To the authors' knowledge, the problem of HDR light field acquisition has only been addressed in [61] and [62]. The authors in [61] apply a spatio-temporal exposure pattern while capturing frames of a light field with a camera array. Four different exposures are captured by neighboring cameras of the camera array. The captured set of frames is used to estimate local point spread functions (PSF) and depth maps, and to segment the scene into regions sharing PSFs of similar shapes at similar depths. The local PSF are also used for motion-deblurring of all image segments at each exposure. The authors in [62] propose an HDR light field acquisition method based on multiple plenoptic captures with varying exposures. The method recovers the highly saturated pixels using the lower exposure captures by computing a weighted low rank approximation of the matrix formed by all the aligned views, with weights defined as a function of the pixels saturation level. The non-saturated pixels are associated to the maximum weight indicating a high confidence, while the fully saturated pixels having low confidence, hence a null weight.

In contrast with the above methods, we propose an acquisi-

$$\underbrace{\begin{bmatrix} \boldsymbol{I}^{(1)} \\ \boldsymbol{I}^{(2)} \\ \vdots \\ \boldsymbol{I}^{(n_\mathrm{s})} \end{bmatrix}}_{\mathbf{I}} = \underbrace{\begin{bmatrix} \boldsymbol{\phi}^{(1,\mathrm{R},1)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{R},1)} & \boldsymbol{\phi}^{(1,\mathrm{G},1)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{G},1)} & \boldsymbol{\phi}^{(1,\mathrm{B},1)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{B},1)} \\ \boldsymbol{\phi}^{(1,\mathrm{R},2)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{R},2)} & \boldsymbol{\phi}^{(1,\mathrm{G},2)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{G},2)} & \boldsymbol{\phi}^{(1,\mathrm{B},2)} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{B},2)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\phi}^{(1,\mathrm{R},n_\mathrm{s})} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{R},n_\mathrm{s})} & \boldsymbol{\phi}^{(1,\mathrm{G},n_\mathrm{s})} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{G},n_\mathrm{s})} & \boldsymbol{\phi}^{(1,\mathrm{B},n_\mathrm{s})} & \dots & \boldsymbol{\phi}^{(\nu,\mathrm{B},n_\mathrm{s})} \end{bmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{bmatrix} \boldsymbol{L}^{(1,\mathrm{R})} \\ \vdots \\ \boldsymbol{L}^{(\nu,\mathrm{R})} \\ \boldsymbol{L}^{(1,\mathrm{G})} \\ \vdots \\ \boldsymbol{L}^{(\nu,\mathrm{G})} \\ \boldsymbol{L}^{(1,\mathrm{B})} \\ \vdots \\ \boldsymbol{L}^{(\nu,\mathrm{B})} \end{bmatrix}}_{\mathbf{L}} , \quad (1)$$

tion design that considers the combination of spatially varying ISO photography principles with those of light field acquisition on a single 2D sensor using coded masks. Our proposed unified framework reconstructs the HDR light field from a single 2D coded projection on the sensor by solving HDR reconstruction, demosaicing, and light field recovery problems simultaneously. The reconstruction problem is placed in a compressive sensing framework that we briefly recall below.

d) Compressive sensing: The compressive sensing theory [63] relies on the assumption that the signal is sparse (or *compressible*) in some transform domain like wavelets, discrete cosine transform (DCT), or dictionaries learned from large datasets. This framework has been already considered for extending the dynamic range of video imaging from interlaced exposures in [64], where sparse coding with learned dictionaries is used for jointly solving deinterlacing and alleviating the noise that arise in interlaced video imaging with different exposures.

In the context of compressed acquisition of light fields using coded masks, original light fields can be restored by solving a basis pursuit denoising (BPDN) problem [65], given an overcomplete dictionary as described in [14], [15].

III. Proposed framework

A. Compressed LDR acquisition of light fields

a) Light field parameterization: In the following sections we will denote the light field as $L(x, y, u, v, \lambda)$ describing the radiant flux in the wavelength λ along the direction passing through two points (x, y) and (u, v) located on two reference planes. Here, (x, y) and (u, v) are usually called spatial (pixel) and angular (view) coordinates respectively. For the sake of simplicity, we assume that the light field L is sampled in three color channels (red, green and blue), and according to a regular rectangular grid with angular resolution $\nu = \nu_u \nu_v$ and spatial resolution $n = n_x n_y$. We then denote by $\mathbf{L} \in \mathbb{R}^{3\nu n}$ the column vector which represents the sampled and vectorized version of L, with the following arrangement $\mathbf{L} = \{\boldsymbol{L}^{(j,\mathrm{c})}\}_{1\leq j\leq \nu, \mathrm{c}\in\{\mathrm{R,G,B}\}}$, where $\boldsymbol{L}^{(j,\mathrm{c})} \in \mathbb{R}^n$ denotes the λ_c-color component of the j-th sampled viewpoint. In other words, $\mathbf{L}$ represents the collection of ν viewpoints, each of which is of the size $n_x \times n_y$ pixels.

b) Mask-based compressive light field sensing: We consider a mask-based light field camera architecture equipped with standard optical components: a main lens, an amplitude coded mask (also called attenuation mask) placed between the main lens and the camera sensor, and eventually a color filter array (CFA) mounted directly on the sensor. The architecture is depicted in Fig. 1. The colour coded mask implements the sensing matrix with transmittance or attenuation factors that depend on the wavelength (or colour component) of the light ray. Casting this operation in a compressed sensing framework, this mask performs a spectro-angular compression of the light field. However, classical sensors have in-built color filter arrays (CFA) which can be seen as equivalent to a mask that samples the color of each measurement recorded on the sensor. The actual equivalent sensing mask is the combination of the color coded sensing mask and of the sensor CFA.

The coded mask which is separated from the sensor by a small distance is a random mask as in [14], [15], [57]. For a given mask and CFA pattern, incoming light rays are therefore filtered by the implemented coded mask and the CFA before they reach the photo-sensitive layer of the sensor. This forms a 2D image on the sensor which represents an optical coded-projection of the scene. Multiple coded projections can be obtained by changing the mask and/or CFA pattern. In practice, one can simply consider mechanical system which allows slightly translating the physical mask around a reference position to create different mask patterns, and thus different coded projections of the same scene.

Based on the compressed sensing paradigm for single-sensor light field photography as in [14], [15], the (optical) imaging equation for the light field $\mathbf{L}$ is expressed as:

$$\mathbf{I} = \boldsymbol{\Phi}\mathbf{L} \ , \quad (2)$$

where $\mathbf{I} \in \mathbb{R}^{n_\mathrm{s} n}$ denotes the vector gathering the captured coded projections of $\mathbf{L}$, and n_s denotes the number of captured images (also called number of acquisitions or number of shots). Since we assume a CFA-equipped sensor, we utilize the dual-mask camera model for incorporating an attenuation mask and a CFA that was proposed in [66]. Here, we assume that the sensor is of the same size as the monochromatic angular image $\boldsymbol{L}^{(j,\mathrm{c})}$ (i.e. $n_x \times n_y$ pixels). The matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n_\mathrm{s} n \times 3\nu n}$ is called the "*coded projection matrix*" describes the filtering and the integration of light rays on each pixel of the sensor. This matrix is in fact the product of two matrices [67], a sensing matrix representing the color coded mask and another matrix corresponding to the CFA integrated within the sensor. The product is performed as an element-wise multiplication in

the spatial domain. In other words, each pixel in the CFA is multiplied by all the angular values in $\mathbf{\Phi}$ corresponding to the pixel, since each CFA pixel affects all the incoming light rays to the corresponding pixel; see [66] for a detailed formulation.

The expanded form of the above equation is given in (1), on top of the page, where each sub-matrix $\Phi^{(j,c,s)} \in \mathbb{R}^{n\times n}$ is diagonal (c.f. [66], Section V.B for detailed explanations on $\Phi^{(j,c,s)}$). From a compressive sensing perspective, the images $\mathbf{I}$, which are spectro-angular compressed versions of the original light field $\mathbf{L}$, can be seen as the set of measurements with respect to the sensing operator $\mathbf{\Phi}$. We define the sampling rate (with respect to $\mathbf{\Phi}$) as the ratio between the number of recorded measurements and the total number of measurements in the original data:

$$R_{\text{sampling}} = \frac{n_{\text{s}} n}{3n\nu} = \frac{n_{\text{s}}}{3\nu} \ .$$

For example, given a color (full RGB) light field with $\nu = 5\times 5$ views, the sampling rate of a two-shot acquisition (i.e. $n_{\text{s}} = 2$) is about 2.67%.

c) Noise model for multi-ISO sensor: To describe the noise at the sensor level where each pixel can have different gain values (corresponding to different ISO settings), we first consider the compound-Gaussian noise model of [26]. According to [26], the electronic signal before the amplification stage on each pixel is corrupted by mixed Poisson-Gaussian noise accounting for photon and dark current shot noise, readout noise, and spatial noise. Given the coded projections $\mathbf{I}$ describing the radiant power incident on the sensor and the exposure time t, the pre-amplification measurements $\widetilde{\mathbf{I}}$ (in electron) can be expressed as:

$$\widetilde{\mathbf{I}} = \mathcal{P}\big[t(\mathbf{I} + \mathbf{I}_{\text{dark}})\big] + \mathbf{n}_{\text{read}} \ , \tag{3}$$

where $\mathcal{P}$ denotes the Poisson noise generator (i.e., $\mathcal{P}(\Lambda)$ is a Poisson variable of parameter Λ), $\mathbf{I}_{\text{dark}} \in \mathbb{R}^n$ represents the (eventually) spatially-varying dark current level (in electrons per second) of the sensor, and $\mathbf{n}_{\text{read}}$ which follows the Gaussian distribution $\mathcal{N}(0, \sigma^2_{\text{read}})$ denotes the electronic read-out noise (in electrons). For the sake of simplicity, we assume that, on average, one photon generates one photo-electron in (3) (i.e. the quantum efficiency is assumed to be 100%). In practice, sensors on consumer cameras possess much lower quantum efficiency coefficient (e.g. front-illuminated sensors are around 50-60$\%$ quantum-efficient [68]).

While raw sensor image intensity is often assumed to be ideally linear with the number of generated electrons; in reality, non-linearities may occur at different stages during the procedure of conversion from electrons to digital units (DUs) [69], [70]. More interestingly, there are image sensors which are intentionally designed to obtain a theoretical non-linear response of the form of a power function [71], [72]. In the context of HDR imaging, these sensors enable nonuniform sampling (quantization) at the analog-to-digital converter (ADC) level and thus allow efficient use of ADC bit-depth to encode specific ranges of light intensity. For the sake of simplicity, let us assume in the scope of this work that the sensor response follows a simple power law $(\cdot)^{\frac{1}{\gamma}}$ (where the value γ depends on the sensor characteristics). By combining the noise model in [26] and the non-linear sensor response function, one can express the digitized output images as follows:

$$\widetilde{\mathbf{I}}_{\mathbf{H}} = \min\left\{I_{\max}, (\mathbf{H}\widetilde{\mathbf{I}} + \mathbf{I}_0)^{\frac{1}{\gamma}} + \mathbf{n}_{\text{quan}}\right\} \ , \tag{4}$$

where $\mathbf{H}$ is a diagonal matrix called the "*gain matrix*" (in DU per electron), $\mathbf{I}_0$ and $I_{\max}$ (in DU) are respectively the camera offset (i.e., the camera's black level) and the maximal pixel intensity, and $\mathbf{n}_{\text{quan}}$ (in DU) which follows the $[-1/2, 1/2]$-uniform distribution represents the noise component induced by the quantization process at the ADC. We note that each diagonal entries of $\mathbf{H}$ is the gain of the corresponding pixel and equals to the product between the camera's electron-to-DU conversion rate and the amplification factor associated with the pixel ISO value. To simplify the pipeline, we assume that each pixel's amplification factor and its ISO setting are proportional, and we consider an amplification factor of 1 for ISO 100. Accordingly, if the pixel's capacity (also *full well depth*) is FWD (in electrons) and the ADC resolution is n_{bits} (in bits), the theoretical gain can be computed as follows:

$$h = \frac{\text{ISO}}{100} \times \frac{2^{n_{\text{bits}}} - 1}{\text{FWD}} \ ,$$

where ISO denotes the associated ISO value of this pixel for a given acquisition setting. Here, $\frac{\text{ISO}}{100}$ and $\frac{2^{n_{\text{bits}}}-1}{\text{FWD}}$ represent the amplification factor and the electron-to-DU conversion rate respectively. To summarize, the imaging chain is depicted in Fig. 2.

d) Calibration of multi-ISO sensors: To handle sensors with different gain settings, we calibrate the readout noise, for each gain/ISO setting. Each digital pixel value is independently transformed to an estimate of the number of photoelectrons reaching the sensor per unit time. To compensate for the readout noise and dark current noise bias, we subtract a bias frame from each observation I_j as

$$\widehat{\mathbf{I}} = \frac{1}{t}\mathbf{H}^{-1}\left[(\widetilde{\mathbf{I}}_{\mathbf{H}})^{\gamma} - \mathbf{I}_{\text{bias}}\right] \ , \tag{5}$$

where each entry $\widehat{\mathbf{I}}_i$ of $\widehat{\mathbf{I}}$ denotes the approximation of the incident radiant power representing the actual observations from which we try to recover the entire HDR light field. The bias frame $\mathbf{I}_{\text{bias}}$ is computed as the average of a large set of black images captured with the same camera settings as the observations but with the lens covered, so that no photons reach the sensor.

The variance of these observations can be estimated from the black frame as

$$\hat{\sigma}^2_{\widehat{\mathbf{I}}_i} = {\mathbf{H}_i}^2 t \widehat{\mathbf{I}}_i + \hat{\sigma}^2_{\text{read}}(\mathbf{H}_i) + \hat{\sigma}^2_{\text{quan}}, \tag{6}$$

and variance can be used as a measure of uncertainty of our input observation $\widehat{\mathbf{I}}_i$ by using the following:

$$W_i = \frac{\mathbf{H}_i^2 t^2 . [\mathbf{I}_i < I_{\max}(\mathbf{H}_i)]}{\hat{\sigma}^2_{\widehat{\mathbf{I}}_i}}. \tag{7}$$

Reconstructing the light field $\mathbf{L}$ from its compressed and recorded projections $\widehat{\mathbf{I}}_i$ requires not only the estimation of the original viewpoints, i.e., recovering the parallax, but also the full RGB color values of each pixel (i.e., performing demosaicing), in the presence of sensor noise.

In the following, we describe two different reconstruction methods. In the first method, the data is captured with multi-shot coded projections using a single ISO for each shot; i.e.

the captured coded images are LDR. We then reconstruct a light field corresponding to each ISO separately using an LDR dictionary. The reconstructed LDR light fields are merged view-per-view using a 2D HDR image reconstruction method from multiple exposures [24], to reconstruct the HDR light field. This method is taken as a baseline for comparisons to our approach since both compressive LDR light field reconstruction, as well as HDR image fusion algorithms are well-researched topics. The second approach jointly performs light field reconstruction with parallax, color demosaicing and HDR reconstruction, which is the main contribution in this paper.

B. Individual Light Field Reconstruction and Fusion

For the purpose of comparisons in the evaluation of our approach, this section describes a variant of the above light field capture and reconstruction method. In this approach, we recover the HDR light field in two steps. First, we capture a set of differently exposed LDR light fields, each reconstructed from a 2D coded projection image but without the spatially varying gain, i.e. as in [14], [15]. The exposure variation is captured by varying the ISO for the full sensor between the different exposures. In a second step, the output HDR light field can then be recovered by merging the LDR light fields using any standard method for HDR fusion, for an overview see [73].

The LDR light field reconstruction is carried out in the following way. Let $\widetilde{\mathbf{I}}_{h_i}$ be the vector formed by the recorded measurements of the light field at a given ISO setting, h_i. The recorded measurements are first calibrated by subtracting a black frame as $\widehat{\mathbf{I}}_{h_i} = \left[(\widetilde{\mathbf{I}}_{h_i})^{\gamma} - \mathbf{I}_{\text{bias}}\right]$ as explained in Section III-A. We then solve for the sparse coefficients of the light field as follows

$$\boldsymbol{\alpha}^*_{h_i} = \arg\min_{\boldsymbol{\alpha}} \left\| \widehat{\mathbf{I}}_{h_i} - \boldsymbol{\Phi}\mathbf{D}_{\text{LDR}}\boldsymbol{\alpha} \right\| + \mu||\boldsymbol{\alpha}||_1, \qquad (8)$$

using one dictionary $\mathbf{D}_{\text{LDR}}$ learned from LDR light fields [66]. The LDR light field at each ISO can thus be reconstructed as $\mathbf{L}_{h_i} = \mathbf{D}_{\text{LDR}}\boldsymbol{\alpha}^*_{h_i}$.

Once the light fields have been reconstructed at the different ISOs $h_i, i = 1 \dots k$, we use the multiple exposure method presented by Debevec and Malik [24] to recover the HDR light field. For the examples used in our evaluations, see Table VII.

As another type of reference for the evaluation, we also use HDR light fields captured as described above but where the exposure time is varied between the exposures instead of the ISO. The LDR light fields are recovered as described in (8), and the HDR light field is computed using the method in [24]. The results are shown in Table IV as a basis for comparison.

C. Joint Light Field Reconstruction and HDR Recovery

To jointly recover the spatio-angular information of the light field, as well as performing demosaicing and HDR recovery, we solve the following dictionary-based minimization per $4D$ patch

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \left\| \mathbf{W}\left(\boldsymbol{\Phi}\mathbf{D}_{\text{HDR}}\boldsymbol{\alpha} - \widehat{\mathbf{I}}\right) \right\|_2^2 + \mu\|\boldsymbol{\alpha}\|_1, \qquad (9)$$

where $\widehat{\mathbf{I}}$ denotes the vector containing the calibrated observations obtained from (5) based on the noise model introduced in (4). The diagonal matrix $\mathbf{W}$ contains the measurement uncertainty for each pixel based on the saturation and noise estimation obtained from (7). Similar to previous work [14], [15], we consider 4D light field patches in order to preserve inter-view angular consistency, as well as efficient reconstruction of the parallax.

We consider a prior on the light field $\mathbf{L}$, by assuming that $\mathbf{L}$ is sparse when it is represented by a suitable dictionary $\mathbf{D}_{\text{HDR}}$. The sparse vector representing the light field $\mathbf{L}$ is denoted as $\boldsymbol{\alpha}$. Since the reference light field contains high dynamic range values, the trained dictionary should as well contain HDR atoms. Therefore, we train the dictionary from a set of HDR light field data sets created in this paper. In practice, due to the large size of light field data, we train the dictionary $\mathbf{D}_{\text{HDR}}$ using small patches of size $5 \times 5 \times 9 \times 9 \times 3$ obtained by dividing the light field of 5×5 views into patches of spatial dimension 9×9 to reduce the learning time of dictionaries; the last dimension defines the spectral domain and in this paper we assume RGB light fields. The reconstruction is performed on light field patches of the same size as the dictionary atoms, and the patches of the final reconstructed light field are computed by averaging pixels on overlapped patches. In the experiments, we used an overlapping distance of 5 pixels for spatial patches of size 9×9. The dictionary $\mathbf{D}_{\text{HDR}}$ is trained on light field patches using the K-SVD algorithm [74][1].

The dictionary-based sparse recovery problem in (9) can be solved with various solvers, e.g. OMP [75], LARS [76], and ISTA [77], as used in [14], [15], [66] with different acquisition schemes. In the experiments reported here, we used the ADMM [78] method to solve the l_1 constrained minimization problem in (9).

Fig. 3: Thumbnails of the tone mapped training set from the real HDR light fields: from left to right and top to bottom, *Bike Closeup 2*, *Bike Close up 4*, *Building Closeup 4*, *Window close up 2*, *Window close up4*, *Outdoor 2*, *Dude 2*, *Glass 2*, *Car close up3*.

[1]The K-SVD toolbox is available at http://www.cs.technion.ac.il/ ronrubin/software.html

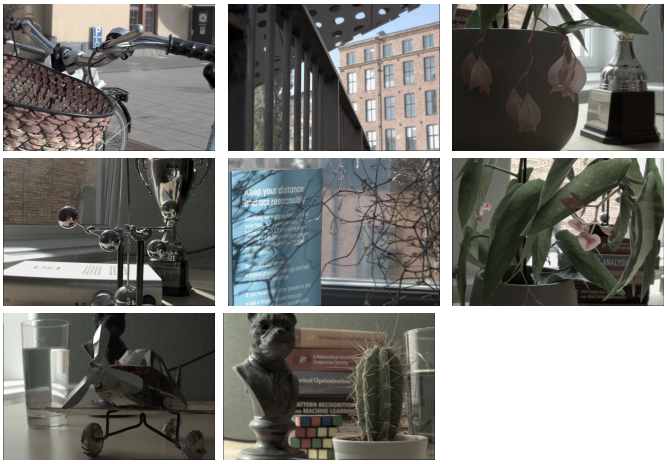


Fig. 4: Thumbnails of real HDR light fields (tone mapped) created from multiple LDR captures used in the test set: from left to right and top to bottom, *Bike*, *Building*, *Trophy*, *Pendulum*, *Window*, *Begonia*, *Plane*, *Cactus*.

Fig. 5: Thumbnails of our tone mapped synthetic HDR light fields: from left to right and top to bottom, *Barbershop*, *Cars*, *Classroom*, *Cosmos*, *Pabellon*. We used *Classroom* and *Pabellon* as the test set, while the rest were used for training, with a baseline of 1, 3, 5 for each.

IV. Results

A. Data sets

One of the key challenges for research on the acquisition and reconstruction of HDR light field is the availability of data sets. To the best of our knowledge there is a lack of HDR light field data sets, except for the one in [79] which contains six static light-fields captured by a camera mounted on two linear axes using exposure bracketing at each view point. This data set however has a large baseline between the neighboring views and is not suitable for evaluating a single sensor light field camera. We, therefore, created the following two publicly available data sets[2]. Each light field in these data sets is accompanied with a histogram to show the dynamic range.

Natural HDR light field data set: To create this data set, we used exposure bracketing to capture 5 exposures, that are 1EV apart, of static scenes using the Lytro Illum camera. To decode the captured light fields we used the toolbox from [80], [81]. Multi-exposed light fields are then merged using the fusion method in [24] to create natural HDR light fields. This data set consists of light field images of a variety of environments from outdoor to indoor with different illuminations. We have used 17 scenes with an angular resolution of 5×5 views and a spatial resolution of 505×354, which was divided into 9 light fields for training an HDR dictionary $\mathbf{D}_{\mathrm{HDR}}$ (see Fig. 3), and 8 light fields for testing (see Fig. 4).

Synthetic HDR light field data set: We have created a synthetic light field data set using *Blender* [82] that contains 5 scenes, which we call *Barbershop*, *Cars*, *Cosmos*, *Classroom*, and *Pabellon*. Each light field contains 5×5 views with a resolution of 960×540. Each scene has been rendered with 5 different baselines. Therefore, in total, we have created 25 rendered HDR light fields. The baseline here is defined as the distance in millimeters between the neighboring virtual cameras that reside on a plane in Blender. Fig. 5 shows central views of these light fields. Three light fields are used for training an HDR dictionary, $\mathbf{D}_{\mathrm{HDR}}$, namely *Barbershop*, *Cars*, and *Cosmos*, each with a baseline of 1 and 3. We used the *Classroom* light field with baselines of 1, 3 and 5 and the *Pabellon* light field with baselines of 1, 3, 5 and 20 as the test set to evaluate our method, see Section IV-C.

B. Evaluation Scenarios

To evaluate the proposed acquisition schemes and the joint spatio-angular-HDR reconstruction we compared different configurations for the spatially varying gain/ISO patterns using different ISO values. As shown in Fig. 1, we compared row-based, per-pixel random, and block-based gain patterns. We used the following three sets of 4 ISO values in a single capture: (100, 400, 800, 1600), (100, 800, 1600, 3200), and (100, 1600, 3200, 6400). We also considered a dual ISO pattern with the following ISO values (100, 1600). The exposure time for each shot is set as $t = 1/240$ second.

We compared the proposed joint multi-ISO HDR light field acquisition scheme to the two reference methods explained in Sec. III-B, in which HDR light fields are recovered using standard exposure fusion, [24], from a set of different LDR light field exposures captured by varying the global sensor ISO and the exposure time. We have also compared our joint reconstruction with reconstruction in two-steps, where an LDR dictionary is used for recovering an LDR light field, followed by a deep learning approach for estimating the HDR light field [25] from the obtained LDR light field.

In the experiments described below, the metrics used for evaluating the HDR reconstruction quality are PU2-PSNR and PU2-SSIM [83], as well as HDR-VDP2 [84]. To calculate the HDR-VDP2 we have set the following display parameters: Display resolution is set to $[1920, 1200]$ for a display size of 24 inches and a viewing distance of 0.5 meter. For natural light field scenes, *Window*, *Begonia*, *Cactus*, *Plane*, *Bike*, *Trophy*, and *Pendulum*, we have used peak luminance values of $[1000, 1000, 1000, 700, 5000, 7000, 1000, 1000]$ in $\mathrm{cd/m^2}$, respectively. For the synthetic data sets, *Classroom* and *Pabellon*, we used peak luminance values of 1000 and 700 in $\mathrm{cd/m^2}$, respectively.

C. HDR light field reconstruction results

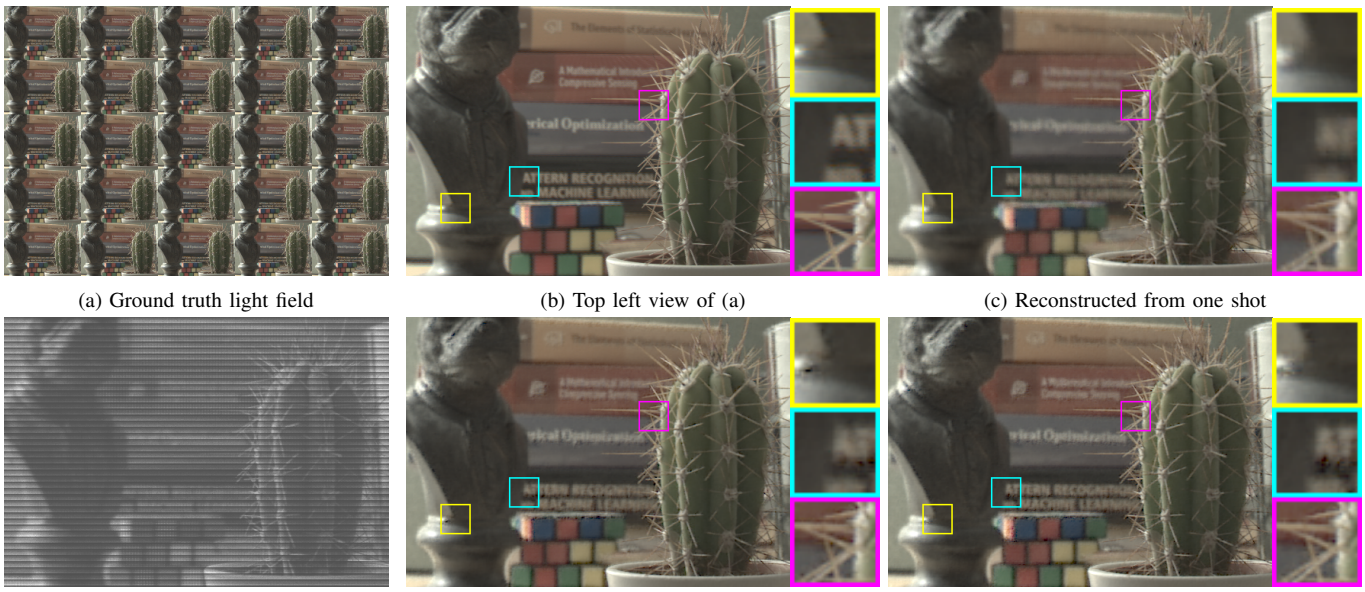
To evaluate the reconstruction results, we trained two HDR dictionaries: one trained on synthetic HDR light fields, and

[2]https://computergraphics.on.liu.se/hdrlf/

CFA Sensor with $1, 2, 3$ shots and with 4 ISO values: $100, 400, 800, 1600$. Joint Reconstruction			
ISO Pattern	PU2 PSNR	PU2 SSIM	HDR VDP2
Block-based	29.41 ; 30.29 ; **30.49**	0.9133 ; 0.9400 ; **0.9422**	55.21 ; 57.93 ; **58.28**
Row-based	31.55 ; 32.73 ; **32.79**	0.9106 ; 0.9407 ; **0.9444**	55.85 ; 58.43 ; **58.70**
Random ISO	31.09 ; **32.84** ; 32.38	0.9072 ; 0.9409 ; **0.9445**	55.63 ; 58.46 ; **58.79**

Individual ISO reconstruction with 4 shots with ISOs: 100, 400, 800, 1600 with exposure time: 0.004		
PU2 PSNR	PU2 SSIM	HDR VDP2
31.25	0.9142	54.18

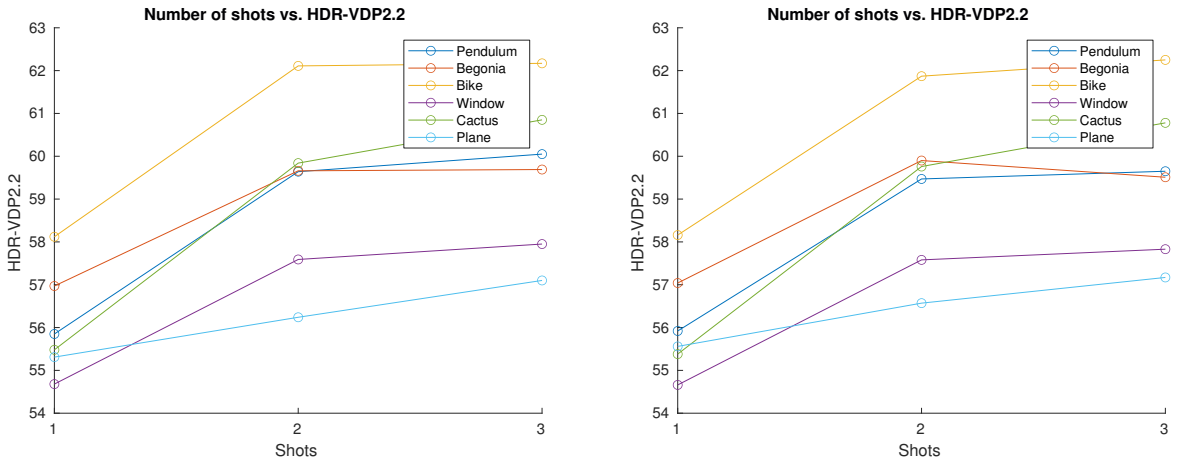
TABLE I: Results of the proposed joint reconstruction algorithm in comparison with individual ISO reconstructions averaged over 8 real light fields in the test set. The acquisition scheme assumes a CFA sensor, four ISO values ($100, 400, 800$ and 1600) with three different ISO patterns: block-based, row-based and random. Results for each light field are shown in Table VII in the Appendix. The regularization strength, μ in (9), has been set to 0.0001. These results have been achieved with $1, 2$ and 3 acquisitions (shots). The exposure time has been set to 0.004 seconds. The stopping criterion of the ADMM algorithm is set to 2^{-18}. Moreover, we set $\gamma = 2.2$.



(a) Ground truth light field (b) Top left view of (a) (c) Reconstructed from one shot

(d) Captured LDR image (e) Reconstructed from two shots (f) Reconstructed from three shots

Fig. 6: LDR sensor image and the reconstruction results of the *Cactus* light field assuming a CFA sensor with four ISO values (100, 400, 800 and 1600) in a row-based pattern using one, two, and three shots.



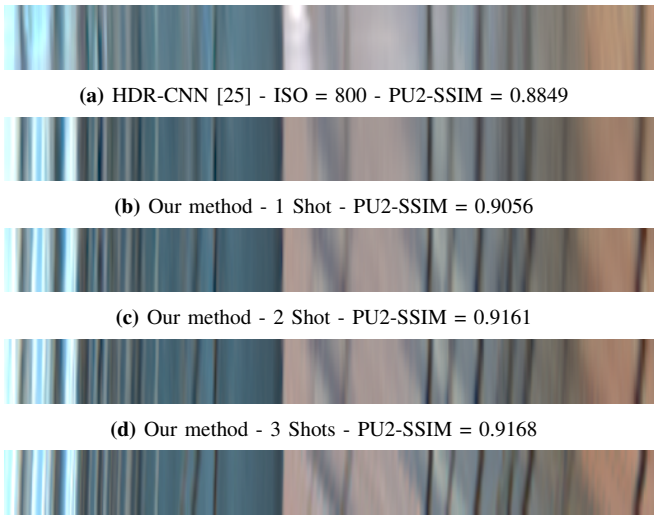
(a) Random (b) Row-wise

Fig. 7: The effect of number of shots on the reconstruction quality with two gain patterns: *Random* and *row-based*.

one on real HDR light fields (see Sec. IV-A) using the K-SVD algorithm [85], with an over-completeness factor of 2.7. We divided the light field into patches of dimension $5\times5\times9\times9\times3$, where 5×5 is the angular dimension, 9×9 the spatial dimension, and 3 for the spectral domain. To train each dictionary, for synthetic and natural light fields, we randomly extracted 300000

HDR-CNN + dictionary learning - ISO values: 100, 400, 800, 1600			
Light Field	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	20.35 ; 25.23 ; 25.22 ; 19.35	0.8542 ; 0.8629 ; 0.8617 ; 0.8373	54.90 ; 55.05 ; 54.61 ; 46.15
Building	20.71 ; 23.79 ; 19.15 ; 8.58	0.8709 ; 0.8785 ; 0.8484 ; 0.5988	53.02 ; 52.31 ; 45.42 ; 41.20
Trophy	19.11 ; 21.69 ; 19.24 ; 23.37	0.8518 ; 0.8935 ; 0.8652 ; 0.8949	47.97 ; 53.34 ; 49.12 ; 48.17
Pendulum	19.93 ; 18.43 ; 19.89 ; 15.95	0.8528 ; 0.8407 ; 0.8511 ; 0.7834	52.28 ; 51.61 ; 49.31 ; 46.78
Window	23.89 ; 27.40 ; 24.90 ; 10.03	0.8054 ; 0.8054 ; 0.7850 ; 0.5174	51.08 ; 50.46 ; 44.09 ; 41.15
Begonia	18.94 ; 26.14 ; 16.94 ; 11.70	0.8738 ; 0.9062 ; 0.8218 ; 0.4750	55.04 ; 54.14 ; 52.81 ; 49.23
Cactus	18.76 ; 21.02 ; 25.55 ; 18.17	0.8413 ; 0.8504 ; 0.8627 ; 0.8149	52.49 ; 52.23 ; 51.98 ; 48.21
Plane	18.48 ; 18.55 ; 16.14 ; 17.78	0.7897 ; 0.8115 ; 0.7695 ; 0.7912	50.68 ; 53.91 ; 52.82 ; 50.16
Average	20.02 ; **21.53** ; 20.88 ; 15.62	0.8425 ; **0.8561** ; 0.8332 ; 0.7141	52.18 ; **52.88** ; 50.02 ; 46.38

TABLE II: Reconstruction with the proposed dictionary learning method and HDR-CNN method [25] for four different gain values: $100, 400, 800, 1600$. The light field reconstruction parameters are the same as the ones used for Table VII.



(a) HDR-CNN [25] - ISO = 800 - PU2-SSIM = 0.8849

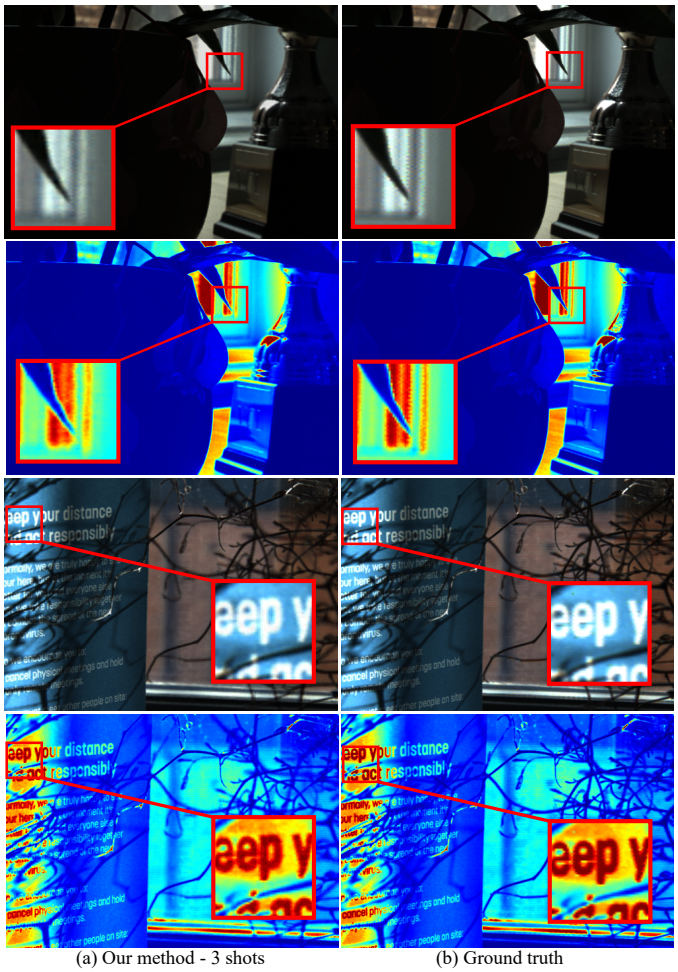
(b) Our method - 1 Shot - PU2-SSIM = 0.9056

(c) Our method - 2 Shot - PU2-SSIM = 0.9161

(d) Our method - 3 Shots - PU2-SSIM = 0.9168

(e) Ground truth

Fig. 8: Comparison of epipolar plane image (EPI) of the *Window* scene. The central row has been selected for display. The EPI is scaled $10\times$ for presentation. Our result is simulated with the row-based pattern with ISO values $= \{100, 400, 800, 1600\}$. The light field reconstruction parameters are similar to Table VII.



(a) Our method - 3 shots (b) Ground truth

Fig. 9: False-color comparison of the HDR values of the *Trophy* and *Window* scenes. The reconstructions by our proposed method are produced with the row-based pattern and 3 shots. The HDR images above the false-color images show the saturated regions of the scene.

patches from the corresponding training set. The spatial patch size, which is set to 9×9 in our experiments, was chosen such that we obtain adequate number of patches for training while keeping the training time tractable. In practice, larger patch sizes lead to a better reconstruction quality, as long as we have an adequate number of patches and that the training time is reasonable. The sparsity parameter for the K-SVD algorithm is set to 10 and the maximum number of training iterations is set to 20. Since we use overlapping patches (in the spatial domain), after reconstructing each light field patch using the ADMM algorithm, we take the average of pixels in overlapping regions to obtain the final result. In the results reported here, we consider half overlapping, i.e., an overlap of 5 pixels for a spatial patch size of 9×9 pixels.

The results from our joint multi-ISO reconstruction are shown in Table I for real light fields. The table contains comparisons to the baseline method described in Section III-B, where 4 shots with 4 ISO values of 100, 400, 800, 1600 are captured and fused for each reference HDR light field. To evaluate how the reconstruction quality varies with the number of input images, we include results for 1 to 3 shots, as well as different ISO patterns. As shown in Table I, random ISO patterns achieve slightly better reconstruction quality as compared to the the other ISO patterns. Increasing the number of acquisitions from 1 shot to 2 shots shows significant improvement, while having 3 shots only slightly affects the reconstruction quality. Visual quality comparisons with respect to the number of shots is presented in Fig. 6, followed by quantitative results in Fig. 7, where we plot the number of shots vs. HDR-VDP2 for each natural light field in the test set. When the number of samples increases, the accuracy of our algorithm for recovering the light field baseline increases as well, see Fig.

CFA Sensor combined with row-based gain patterns (using two ISO values: 100 and 1600) and with 1, 2 and 3 shots			
Light Field	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	28.03 ; **28.80** ; 28.63	0.9235 ; **0.9434** ; 0.9427	57.55 ; 60.93 ; **61.22**
Building	**29.92** ; 29.28 ; 29.22	0.8974 ; 0.9313 ; **0.9331**	54.31 ; 56.05 ; **57.14**
Trophy	29.52 ; 32.03 ; **32.42**	0.9273 ; **0.9367** ; 0.9326	**53.74** ; 52.84 ; 52.01
Pendulum	**25.83** ; 25.65 ; 25.45	0.9027 ; 0.9196 ; **0.9197**	55.61 ; 58.56 ; **58.74**
Window	30.79 ; **32.34** ; 32.28	0.8966 ; 0.9397 ; **0.9484**	53.96 ; 57.38 ; **58.55**
Begonia	**30.91** ; 30.16 ; 29.84	0.9356 ; **0.9490** ; 0.9481	56.67 ; **59.45** ; 59.11
Cactus	31.60 ; 32.68 ; **32.75**	0.9040 ; 0.9439 ; **0.9483**	54.35 ; 58.94 ; **59.56**
Plane	29.15 ; 31.75 ; **32.19**	0.9145 ; 0.9311 ; **0.9313**	55.09 ; 55.28 ; **55.39**
Average	29.47 ; 30.34 ; **30.35**	0.9127 ; 0.9368; **0.9380**	55.16 ; 57.43 ; **57.71**

TABLE III: Reconstruction results of the proposed joint algorithm using a row-based pattern with two gain values and real light fields. The regularization strength, μ in (9) has been set to 0.0001. These results have been achieved with $1, 2$, or 3 acquisitions. The exposure time has been set to 0.041 seconds. The stopping criterion of the ADMM algorithm has been set to 2^{-18}. Moreover, we set $\gamma = 2.2$

CFA Sensor with fusion of $2, 3, 4$ captures at different exposures.			
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	31.46 ; 32.24 ; 32.71	0.927 ; 0.935 ; 0.938	56.99 ; 58.44 ; 58.79
Building	31.84 ; 32.37 ; 32.51	0.905 ; 0.908 ; 0.907	54.56 ; 54.63 ; 54.60
Trophy	35.88 ; 36.11 ; 36.22	0.955 ; 0.957 ; 0.958	56.15 ; 56.30 ; 56.55
Pendulum	30.47 ; 30.72 ; 30.90	0.926 ; 0.925 ; 0.926	54.93 ; 54.70 ; 54.62
Window	30.63 ; 31.03 ; 31.06	0.859 ; 0.871 ; 0.872	52.90 ; 53.09 ; 52.99
Begonia	33.93 ; 34.13 ; 34.20	0.947 ; 0.949 ; 0.950	57.71 ; 57.94 ; 57.91
Cactus	31.75 ; 31.89 ; 31.95	0.898 ; 0.902 ; 0.903	54.26 ; 54.40 ; 54.41
Plane	34.06 ; 34.16 ; 34.21	0.950 ; 0.951 ; 0.952	54.60 ; 54.71 ; 54.60
Average	32.50 ; 32.83 ; **32.97**	0.921 ; 0.925 ; **0.926**	55.26 ; 55.53 ; **55.56**

TABLE IV: Reconstruction results with compressive exposure bracketing. The acquisition scheme assumes a CFA sensor, and 2, 3, and 4 different exposures. The regularization strength, μ in (9), has been set to 0.0001. The relative exposure times used for simulation are $[0.5, 8]$ for 2 shots, $[0.5, 2, 8]$ for 3 shots, and $[0.5, 2, 4, 8]$ for 4 shots. The stopping criterion of the ADMM algorithm has been set to 2^{-18}. Moreover, we set $\gamma = 2.2$

8 and the supplementary video[3]. To show the reconstruction quality with respect to the true HDR values, i.e. without tone mapping, in Fig. 9 we give false-color representations of the reconstructed HDR images for the *Trophy* and *Window* scenes. This figure shows that three shots are adequate for high quality reconstruction.

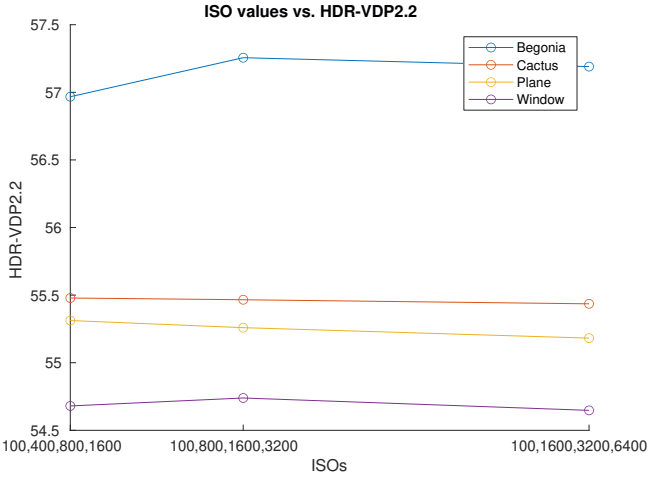


Fig. 10: Reconstruction quality with 3 different ISO values.

Following the introduction of the CNN-based architectures, many methods have been proposed for solving the HDR reconstruction task which have shown promising results. Therefore, we also compare our method with CNN-based approaches. Since, to the author's knowledge, there exists no CNN-based solution that would perform the joint light field reconstruction and HDR recovery, we have compared with a state-of-the-art CNN-based method designed for reconstructing an HDR image from a single LDR image [25], that we applied on each view of the LDR light field reconstructed from compressed measurements. To do this, we first apply the dictionary-based reconstruction method in [66] to recover the full-resolution LDR light field, and in the second step we apply the HDR-CNN method on each view to recover the HDR values. The results are given in Table II for 4 different ISO settings: 100, 400, 800, and 1600. As presented in Fig. 8, this two step reconstruction is not able to recover the baseline between the neighboring views, and the hallucination of saturated areas lacks consistency between the views, which can also be seen in the supplementary video.

Although there is no commercial implementation of a sensor with per-pixel gain, a row-based ISO control of Canon cameras has been utilized in previous work [22]. Therefore, we performed an evaluation using only 2 variations of gain on the sensor with ISO settings $(100, 1600)$ on the row-based pattern to justify the chosen number of ISOs per single capture, see Table III. Comparing the results of Table III with that of Table I, we observe that the higher number of gain variations leads to a better reconstruction quality, as expected.

Table IV gives the results when fusing the reconstructed light fields at different exposures. Note that, when capturing the light field with 4 different exposures, the number of measurements is 4 times higher than what is required by our method with a multi-ISO sensor and with one shot. One should also note that the camera setup for the results of Table IV is different from our multi-ISO experiment. Despite having a fraction of the measurements, our method in most cases can reconstruct the scene better than the method based on capturing and merging 4 differently exposed LDR light fields by varying the exposure

[3]http://clim.inria.fr/research/HDRLF-CS/index.html

CFA Sensor with 3 shots with 4 ISO values: 100, 400, 800, 1600. Noise level: none, low, medium, high.			
Light Fields	HDR VDP2	PU2 PSNR	PU2 SSIM
Classroom - baseline 1	57.41 ; 56.48 ; 55.23 ; 54.03	25.55 ; 25.94 ; 25.82 ; 25.64	0.9088 ; 0.9088 ; 0.9039 ; 0.8964
Classroom - baseline 3	56.79 ; 55.77 ; 55.22 ; 54.07	28.05 ; 28.36 ; 28.19 ; 27.89	0.9312 ; 0.9319 ; 0.9260 ; 0.9172
Classroom - baseline 5	52.48 ; 51.94 ; 52.29 ; 51.77	27.75 ; 27.80 ; 27.62 ; 27.38	0.9205 ; 0.9203 ; 0.9144 ; 0.9057
Pabellon - baseline 1	55.32 ; 55.35 ; 55.25 ; 54.92	27.16 ; 26.97 ; 26.92 ; 26.87	0.9078 ; 0.9022 ; 0.8999 ; 0.8967
Pabellon - baseline 3	55.93 ; 55.49 ; 55.31 ; 55.02	28.34 ; 28.14 ; 28.12 ; 28.10	0.9279 ; 0.9241 ; 0.9230 ; 0.9219
Pabellon - baseline 5	56.01 ; 55.66 ; 55.75 ; 55.82	28.71 ; 28.52 ; 28.50 ; 28.48	0.9317 ; 0.9283 ; 0.9272 ; 0.9261
Pabellon - baseline 20	50.69 ; 50.71 ; 50.41 ; 50.77	28.63 ; 28.42 ; 28.43 ; 28.40	0.8771 ; 0.8709 ; 0.8701 ; 0.8689

TABLE V: Evaluation of our method on the synthetic data set, *Classroom* and *Pabellon*, with four noise levels (none, low, medium, and high) and a random ISO pattern. The baseline between the neighboring cameras increases from 1 to 5 for both data set and to 20 for *Pabellon*. The camera noise parameters are given in Table VI. We set $\mu = 0.001$, $\gamma = 3.0$, $t = 0.041$, and use 3 shots for this experiment. The ADMM stopping criterion is set to 2^{-16}.

time. The same observation holds also when comparing to the baseline method that uses a distinct ISO setting for each shot (see the bottom part of Table I). This can be explained by the fact that the individual reconstruction method with 4 shots proceeds in two steps: the method first separately reconstructs 4 LDR light fields using LDR dictionaries and then fuses the images corresponding to the same viewpoint. First, the separately reconstructed LDR light fields may not be fully consistent in the angular dimension, hence the fusing may blur edges and textures. In addition, the first step of LDR light field reconstruction using LDR dictionaries may not work well in saturated and under-exposed areas, and these errors cannot be well recovered by the fusion of multiple-exposures. In contrast, the joint reconstruction, thanks to the use of a multi-ISO sensor, fully exploits the dynamic range information available in a spatial pixel neighborhood, while taking into account the angular information simultaneously. This allows us to better recover intensity variations between neighbouring pixels.

HDR algorithms are limited by the level of saturation that they can handle. Fig. 10 shows the accuracy as the ISO values used are varied. By increasing the distance between the steps of the ISO parameter, from (100, 400, 800, 1600) to (100, 800, 1600, 3200) and (100, 1600, 3200, 6400), the reconstruction quality remains intact, even though we introduce saturation on pixels that are amplified with higher gain values. This shows the robustness of our algorithm in handling saturated pixels, thanks to the joint recovery approach together with the HDR dictionary. The robustness of our approach in handling saturated pixels is also confirmed when using the synthetic data set, where a higher percentage of pixels are saturated. We present these results in what follows.

Noise level	**Dark current** $\mathbb{E}[\mathbf{I}_{\mathrm{dark}}]$ (electrons/pixel/s)	**Readout noise** σ_{read} (rms electrons/pixel)
Low	0.1	1.1
Medium	1.0	2.0
High	10.	2.8

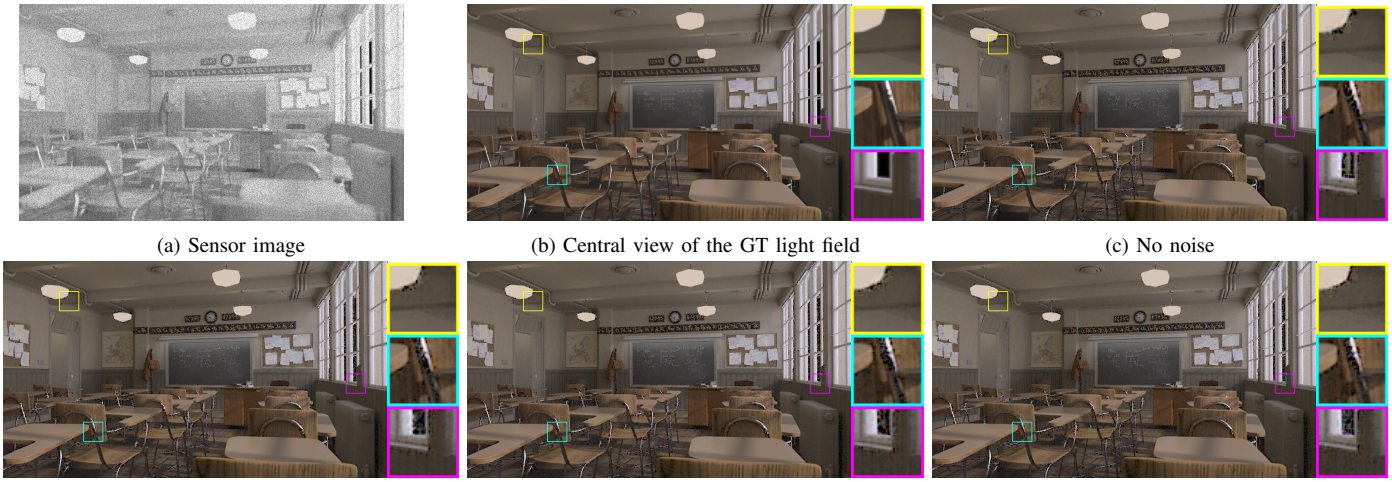
TABLE VI: Camera noise parameters used in the experiments reported in Table V and Fig. 11

In addition to the real HDR light field data set, we have tested our algorithm on the synthetic HDR light fields, introduced in Section IV-A, in order to have control over the level of disparity of light fields and the noise level introduced by different sources during the acquisition process. It should be noted that the synthetic light fields have much larger per-pixel disparity than that of our natural light fields; hence, making this data set particularly challenging for reconstructing the baseline and HDR values. We apply four different levels of noise to the synthetic data set: none, low, medium, and high. More precisely, the noise component before the amplification stage (see Fig. 2) is simulated using (3) with the corresponding parameters of the dark current and the read-out noise shown in Table VI. Afterwards, these noisy measurements are amplified and quantized to produce output sensor images, as depicted in (4). In our simulations, the parameter for non-uniform quantization is set to $\gamma = 3$. For all acquisitions, we set the exposure time of each shot to $t = \frac{1}{240}$ second. We only perform noise analysis on synthetic data set since the natural HDR light field data set already contains some level of camera noise, and adding extra noise to it would not be realistic. As illustrated in Table V, our algorithm works well in both noiseless and noisy simulations, where the presence of different levels of noise only marginally affects the reconstruction quality. To compare the baseline reconstruction quality, the reader is referred to the supplementary video.

Figure 11 demonstrates the visual quality of the reconstructed top-left view of the *Classroom* light field with different noise levels. To evaluate the robustness of our proposed method to larger per-pixel disparities, we have chosen different baselines between the neighboring cameras, changing from 1 to 5 for the *Classroom* and the *Pabellon* scenes. To consider an extreme case, we have also tested baseline 20 for the *Pabellon* scene. Note that such a large baseline is very challenging to reconstruct and is not typically feasible to achieve using a compressive light field camera setup. Yet, we consider it to show the robustness of our reconstruction algorithm. As shown in the table, the reconstruction quality only drops slightly when the disparity increases. The dictionary training was performed on unseen data sets with baselines 1, 3 and 5.

D. Limitations

As all learning-based methods, the recovery works well if the HDR dictionaries have been trained using light fields in the same disparity range as the disparity of the light fields to be processed. Our experiments on synthetic data, where the baseline is controlled by changing the distance between virtual cameras, show that the proposed method is robust to relatively significant increases in baseline, see Table V. Note that the dictionary training for the synthetic data was performed on light fields distinct from the test set with a maximum baseline of 5. Yet we observe that the Pabellon scene used for testing with a baseline of 20 performs well compared to lower baselines.



(a) Sensor image (b) Central view of the GT light field (c) No noise

(d) Low noise level (e) Medium noise level (f) High noise level

Fig. 11: Reconstruction of *Classroom* obtained with random ISO pattern in four noise conditions: no noise, low, medium, and high noise level.

This shows the robustness of our method to changes in camera baseline, although more light fields from a diverse set of scenes are needed to fully confirm this. In addition, the compressed light field acquisition framework works better if the correlation between different views is high, i.e. if the disparity between views is not too large. This is mainly due to the fact that we extract 4D patches from light fields. However, the design considered here, i.e., using one unique sensor for each capture with the aperture size of a normal 2D camera, the disparity of the captured views cannot be too large. Moreover, to recover more views without degrading their quality would require to increase the number of measurements, i.e., the dimension of the sensor or the number of shots for a given sensor size.

V. Conclusions and Future Work

In this paper, we have presented a novel framework for the acquisition and reconstruction of HDR light field images on a single sensor with CFA and per-pixel gain amplification (multi-ISO). The framework solves the problem of demosaicing, HDR, and light field reconstruction jointly in a single step using a learned HDR overcomplete dictionary. We showed that by employing a compressive sensing framework, we can recover a full-resolution HDR light field from a 2D image with acceptable results, which we show quantitatively and qualitatively. We evaluated our framework with different scenarios such as different ISO patterns, varying ISO values, and exposure times. We also investigated the impact of noise on the reconstruction quality by introducing different noise levels in the camera simulation pipeline. Our results show that the proposed framework outperforms the individual light field reconstruction with exposure bracketing when either the ISO setting or the exposure time is changed for each shot.

This is the first exploration of the concept of HDR light field acquisition on a single sensor using coded masks and multiple ISO sensors simultaneously. We have performed a feasibility study, using standard dictionary-based sparse recovery methods. Although this allows us to validate the concept, future work will be dedicated to improving the reconstruction algorithm by considering deep learning techniques that have been shown to be powerful for standard compressive light field acquisition. Another direction for future research is the application of our proposed technique on light field videos. By extending our sensing model to take advantage of the coherence between video frames, similar to [16], one can obtain HDR light field videos using a single sensor equipped with CFA.

VI. Acknowledgments

This work was supported in part by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM), and in part by the Knut and Alice Wallenberg Foundation, the Wallenberg AI, Autonomous Systems and Software Program (WASP), and the strategic research environment ELLIIT.

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. ACM, 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH '96. New York, NY, USA: Association for Computing Machinery, 1996, p. 43–54. [Online]. Available: https://doi.org/10.1145/237170.237200

[3] J. Unger, S. Gustavson, P. Larsson, and A. Ynnerman, "Free form incident light fields," in Proceedings of the Nineteenth Eurographics Conference on Rendering, ser. EGSR '08. Goslar, DEU: Eurographics Association, 2008, p. 1293–1301. [Online]. Available: https://doi.org/10.1111/j.1467-8659.2008.01268.x

[4] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle, "Surface light fields for 3d photography," in Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH '00. USA: ACM Press/Addison-Wesley Publishing Co., 2000, p. 287–296. [Online]. Available: https://doi.org/10.1145/344779.344925

[5] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Handheld Plenoptic Camera," Stanford University, Computer Science Technical Report CSTR 2(11), 2005.

[6] 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 Transactions on Graphics (TOG), vol. 24, no. 3, pp. 765–776, July 2005.

[7] M. Tao, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation for glossy surfaces with light-field cameras," in ECCV Workshop on Light Fields for Computer Vision (L4CV), 2014.

[8] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," ACM Trans. Graph., vol. 25, no. 3, p. 924–934, Jul. 2006. [Online]. Available: https://doi.org/10.1145/1141911.1141976

[9] Z. Zhang, L. Bai, L. Cong, P. Yu, T. Zhang, W. Shi, F. Li, J. Du, and K. Wang, "Imaging volumetric dynamics at high speed in mouse and zebrafish brain with confocal light field microscopy," Nature Biotechnology, vol. 39, no. 1, pp. 74–83, 2021. [Online]. Available: https://doi.org/10.1038/s41587-020-0628-7

[10] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," ACM Trans. Graph., vol. 31, no. 4, Jul. 2012. [Online]. Available: https://doi.org/10.1145/2185520.2185576

[11] J. Yang, M. Everett, C. Buehler, and L. McMillan, "A Real-Time Distributed Light Field Camera," in Proceedings of Eurographics Workshop on Rendering (EGSR), 2002, pp. 77–86.

[12] E. H. Adelson and J. Y. A. Wang, "Single Lens Stereo with a Plenoptic Camera," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 99–106, February 1992.

[13] S. D. Babacan, R. Ansorge, M. Luessi, R. Molina, and A. K. Katsaggelos, "Compressive sensing of light fields," in 2009 16th IEEE International Conference on Image Processing (ICIP), November 2009, pp. 2337–2340.

[14] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," ACM Trans. on Graphics (Proc. of SIGGRAPH), vol. 32, no. 4, pp. 46:1—-46:12, 2013.

[15] E. Miandji, J. Unger, and C. Guillemot, "Multi-shot single sensor light field camera using a color coded mask," in European Signal Processing Conference (EUSIPCO), June 2018, pp. 226–230.

[16] S. Hajisharif, E. Miandji, C. Guillemot, and J. Unger, "Single sensor compressive light field video camera," Computer Graphics Forum, vol. 39, no. 2, pp. 463–474, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13944

[17] C. Ma, Y. Liu, Y. Li, Q. Zhou, X. Wang, and Y. Chang, "A 4-m pixel high dynamic range, low-noise cmos image sensor with low-power counting adc," IEEE Transactions on Electron Devices, vol. 64, pp. 3199–3205, 2017.

[18] T. Willassen, J. Solhusvik, R. Johansson, S. Yaghmai, H. Rhodes, S. Manabe, D. Mao, Z. Lin, D. Yang, O. Cellek, E. Webster, S. Ma, and B. Zhang, "A 1280 x 1080 4.2 µm split-diode pixel hdr sensor in 110 nm bsi cmos process," in 2015 INTERNATIONAL IMAGE SENSOR WORKSHOP. International Image Sensor Society, Jun. 2015.

[19] S. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 472–479 vol.1, 2000.

[20] M. Alghamdi, Q. Fu, A. Thabet, and W. Heidrich, "Reconfigurable snapshot hdr imaging using coded masks and inception network," Vision, p. Modeling and Visualization, 2019.

[21] G. Wan, X. Li, G. Agranov, M. Levoy, and M. Horowitz, "Cmos image sensors with multi-bucket pixels for computational photography," IEEE Journal of Solid-State Circuits, vol. 47, pp. 1031–1042, 2012.

[22] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive dualiso hdr reconstruction," EURASIP Journal on Image and Video Processing, vol. 41, no. 1, 2015.

[23] ——, "HDR Reconstruction for Alternating Gain (ISO) Sensor Readout," in Eurographics 2014 - Short Papers, E. Galin and M. Wand, Eds. The Eurographics Association, 2014.

[24] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH '97. USA: ACM Press/Addison-Wesley Publishing Co., 1997, p. 369–378. [Online]. Available: https://doi.org/10.1145/258734.258884

[25] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "Hdr image reconstruction from a single exposure using deep cnns," ACM Transactions on Graphics (TOG), vol. 36, no. 6, 2017.

[26] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in International Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 553–560.

[27] V. Brajovic and T. Kanade, "A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors," Proceedings of IEEE International Conference on Robotics and Automation, vol. 2, pp. 1638–1643 vol.2, 1996.

[28] S. Nayar and V. Branzoi, "Adaptive dynamic range imaging: optical control of pixel exposures over space and time," Proceedings Ninth IEEE International Conference on Computer Vision, pp. 1168–1175 vol.2, 2003.

[29] S. Kavusi and A. E. Gamal, "Quantitative study of high-dynamic-range image sensor architectures," in Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V, N. Sampat, R. J. Motta, and M. M. Blouke, Eds., vol. 5301. International Society for Optics and Photonics, 2004, pp. 264 –275.

[30] W. Kao, "High dynamic range imaging by fusing multiple raw images and tone reproduction," IEEE Transactions on Consumer Electronics, vol. 54, no. 1, Feb. 2008.

[31] K. Kirk and H. Andersen, "Noise characterization of weighting schemes for combination of multiple exposures," in British Machine Vision Conference (BMVC), 2006, pp. 1129–1138.

[32] K. J., G. S., B. G., and U. J., "Unified hdr reconstruction from raw cfa data," in IEEE International Conference on Computational Photography (ICCP), Apr. 2013.

[33] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," International Journal of Computer Vision, vol. 58, pp. 7–17, 2004.

[34] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," ACM SIGGRAPH 2011 papers, 2011.

[35] Q. Sun, E. Tseng, Q. Fu, W. Heidrich, and F. Heide, "Learning rank-1 diffractive optics for single-shot high dynamic range imaging," IEEE CVPR, 2020.

[36] A. Ito, S. Tambe, K. Mitra, A. C. Sankaranarayanan, and A. Veeraraghavan, "Compressive epsilon photography for post-capture control in digital imaging," ACM Trans. Graph., vol. 33, no. 4, Jul. 2014. [Online]. Available: https://doi.org/10.1145/2601097.2601207

[37] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch, "Optimal hdr reconstruction with linear digital cameras," in CVPR, 2010, pp. 215–222.

[38] J. Hu, O. Gallo, K. Pulli, and X. Sun, "Hdr deghosting: How to deal with saturation ?" in CVPR, 2013.

[39] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust Patch-Based HDR Reconstruction of Dynamic Scenes," ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH Asia 2012), vol. 31, no. 6, pp. 203:1–203:11, 2012.

[40] O. Gallo, A. Troccoli, J. Hu, K. Pulli, and J. Kautz, "Locally non-rigid registration for mobile hdr photography," in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 00, Jun. 2015, pp. 48–55.

[41] T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[42] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," IEEE Signal Processing Letters, vol. 21, no. 9, pp. 1045–1049, Sep. 2014.

[43] C. Lee and E. Y. Lam, "High dynamic range imaging via truncated nuclear norm minimization of low-rank matrix," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2016, pp. 1229–1233.

[44] ——, "Computationally efficient truncated nuclear norm minimization for high dynamic range imaging," IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4145–4157, Sep. 2016.

[45] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Trans. Graph., vol. 36, no. 4, pp. 144:1–144:12, Jul. 2017.

[46] S. Hasinoff and K. Kutulakos, "Confocal stereo," in Computer Vision - ECCV 2006, 2006, pp. 620–634.

[47] M. S. Santos, I. R. Tsang, and N. K. Kalantari, "Single image hdr reconstruction using a cnn with masked features and perceptual loss," ACM Transactions on Graphics (TOG), vol. 39, pp. 80:1 – 80:10, 2020.

[48] Y. Liu, W.-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J. Huang, "Single-image hdr reconstruction by learning to reverse the camera pipeline," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1648–1657, 2020.

[49] C. A. Metzler, I. Hayato, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1372–1382, 2020.

[50] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," Computer Graphics Forum, vol. 35, no. 2, pp. 153–163, 2016.

[51] O. Gallo and P. Sen, "Chapter 3 - stack-based algorithms for HDR capture and reconstruction," in High Dynamic Range Video, F. Dufaux, P. L. Callet, R. K. Mantiuk, and M. Mrak, Eds. Academic Press, 2016, pp. 85 – 119.

[52] S. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive light field sensing," IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4746–4757, December 2012.

[53] E. Miandji, J. Kronander, and J. Unger, "Compressive Image Reconstruction in Reduced Union of Subspaces," Computer Graphics Forum, vol. 34, no. 2, pp. 33–44, May 2015.

[54] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in International Conference on Computer Vision (ICCV), 2017, pp. 4280–4288.

[55] A. K. Vadathya, S. Cholleti, G. Ramajayam, V. Kanchana, and K. Mitra, "Learning light field reconstruction from a single coded image," in Asian Conference on Pattern Recognition (ACPR), 2017.

[56] M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, and P. Turaga, "Compressive light field reconstructions using deep learning," in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jul 2017, pp. 1277–1286.

[57] O. Nabati, D. Mendlovic, and R. Giryes, "Fast and accurate reconstruction of compressed color light field," in 2018 IEEE International Conference on Computational Photography (ICCP), May 2018, pp. 1–11.

[58] Y. Inagaki, Y. Kobayashi, K. Takahashi, T. Fujii, and H. Nagahara, "Learning to capture light fields through a coded aperture camera," in The European Conference on Computer Vision (ECCV), Sep 2018.

[59] M. Guo, J. Hou, J. Jin, J. Chen, and L.-P. Chau, "Deep spatial-angular regularization for compressive light field reconstruction over coded apertures," arXiv:2007.11882, Tech. Rep., 2020.

[60] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," in SIGGRAPH 2007, 2007.

[61] D. C. Schedl, C. Birklbauer, and O. Bimber, "Coded Exposure HDR Light-Field Video Recording," Computer Graphics Forum, 2014.

[62] M. Le Pendu, A. Smolic, and C. Guillemot, "High dynamic range light fields via weighted low rank approximation," in ICIP 2018 - IEEE International Conference on Image Processing. Athens, Greece: IEEE, Oct. 2018, pp. 1–5.

[63] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21–30, Mar 2008.

[64] I. Choi, S.-H. Baek, and M. H. Kim, "Reconstructing interlaced high-dynamic-range video using joint learning," IEEE Transactions on Image Processing, vol. 26, pp. 5353–5366, 2017.

[65] S. Chen, D. Donoho, and M. Saunders, "Atomic Decomposition by Basis Pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.

[66] H. N. Nguyen, E. Miandji, and C. Guillemot, "Multi-mask camera model for compressed acquisition of light fields," IEEE Transactions on Computational Imaging, vol. 7, pp. 191–208, 2021.

[67] H.-N. Nguyen and C. Guillemot, "Color and angular reconstruction of light fields from incomplete-color coded projections," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.

[68] C. Greb, P. Laskey, and K. Schwab. Introduction to digital camera technology. February 02, 2016. [Online]. Available: https://www.leica-microsystems.com/science-lab/introduction-to-digital-camera-technology/

[69] J. Li, Y. fei Guo, and T.-J. Lan, "Study of v/q non-linearity in scientific cmos sensor," International Journal of Grid and Distributed Computing, vol. 9, pp. 289–298, 2016.

[70] F. Wang and A. Theuwissen, "Linearity analysis of a cmos image sensor," Electronic Imaging, vol. 2017, no. 11, pp. 84–90, 2017.

[71] Seogheon Ham, Yonghee Lee, Wunki Jung, Seunghyun Lim, Kwisung Yoo, Youngcheol Chae, Jihyun Cho, Dongmyung Lee, and Gunhee Han, "Cmos image sensor with analog gamma correction using nonlinear single-slope adc," in 2006 IEEE International Symposium on Circuits and Systems, 2006, pp. 4 pp.–3581.

[72] Y. Cao, X. Pan, X. Zhao, and H. Wu, "An analog gamma correction scheme for high dynamic range cmos logarithmic image sensors," Sensors, vol. 14, no. 12, pp. 24 132–24 145, 2014.

[73] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting, 2010.

[74] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcompletedictionaries for sparse representation," IEEE Transactions on Signal Processing, vol. 54, no. 11, Nov. 2006.

[75] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in 27th Asilomar Conference on Signals, Systems and Computers, vol. 1, 1993, pp. 40–44.

[76] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," Annals of Statistics, vol. 32, no. 2, pp. 407–499, 2004.

[77] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

[78] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, ""distributed optimization and statistical learning via the alternating direction method of multiplier," Foundations and Trends in Machine Learning, vol. 3, no. 1, pp. 1–122, 2011.

[79] M. K. Gul, T. Wolf, M. Bätz, M. Ziegler, and J. Keinert, "A high-resolution high dynamic range light-field dataset with an application to view synthesis and tone-mapping," 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–6, 2020.

[80] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. IEEE, Jun 2013.

[81] ——, "Linear volumetric focus for light field cameras," ACM Transactions on Graphics (TOG), vol. 34, no. 2, Feb. 2015.

[82] "Blender website," https://www.blender.org/.

[83] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full dynamic range images," in Human Vision and Electronic Imaging XIII, ser. Proceedings of SPIE, San Jose, USA, January 2008, pp. 6806–10.

[84] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," ACM Transactions on Graphics (Proc. of SIGGRAPH'11), vol. 30, no. 4, 2011.

[85] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. on Signal Processing, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

Ehsan Miandji received his B.Sc. degree from Azad University, Tehran, Iran in 2008, and his M.Sc. and Ph.D. in computer graphics and image processing from Linkoping University, Sweden, in 2012 and 2018, respectively. He pursued post-doctoral research at INRIA, Rennes, France in 2020. Currently, he is a post-doctoral research associate at the Department of Science and Technology, Linkoping University. Dr. Miandji has published peer-reviewed journal papers in compressive light field photography, compression and compressed sensing of visual data, photorealistic image synthesis, as well as theoretical work on compressed sensing. He was an IPC member for Eurographics short papers program in 2020 and 2021. He is a member of IEEE, ACM SIGGRAPH, and Eurographics.

Hoai-Nam Nguyen received the B.S. degree in applied mathematics from Université Paris Descartes in 2011, the M.S. degree in mathematics, computer vision and machine learning from ENS Cachan in 2013, and the Ph.D. degree in signal processing and telecommunication from Université de Rennes 1 in 2017. Since 2018, he has been a Research Engineer with INRIA Rennes – Bretagne Atlantique. He is currently a member of the Sirocco project-team leaded by Christine Guillemot. His research interests include image formation modeling, compressed sensing and reconstruction, convex and non convex optimization, generic algorithms for inverse problems and their applications in light field image processing.

Saghi Hajisharif is currently a post-doctoral researcher at Linköping University, Sweden. She is part of the visual computing laboratory led by Jonas Unger at the department of science and technology. She received her B.Sc. degree in computer science from the Amirkabir University of Technology, Tehran, Iran in 2009, and her M.Sc. and Ph.D. in computer graphics and image processing from Linkoping University, Sweden, in 2013 and 2020, respectively. In 2015, she was a visiting researcher at the Institute for Creative Technologies (ICT) at the University of Southern California (USC). Her research interests are computer graphics and image processing focusing on capturing and reconstruction techniques for high dynamic range imaging, light field photography, and high dimensional visual data using sparse modeling.

Jonas Unger is a professor in computer graphics at Linköping University, Sweden. Unger received MSc and PhD degrees in media technology from Linköping University in 2003 and 2009 respectively, became a docent in 2015, and has been a visiting researcher at the University of Southern California. Since 2009, he is leading the Computer Graphics and Image Processing group at Linköping University with a focus on research and development of new theory and technology for computational imaging fusing computer graphics, vision and sensors with human perception and machine learning. With a strong foundation in theoretically oriented research, Unger and his group are working closely with industry to develop state-of-the-art methods for 3D scene reconstruction, photo-realistic image synthesis and digitization of optical material properties, heart surgery capturing, perceptual displays, and generation of data for autonomous systems in applications domains such as self-driving vehicles and medical imaging.

Christine Guillemot is an IEEE fellow and Director of Research at INRIA. She holds a Ph.D. degree from ENST (Ecole Nationale Superieure des Telecommunications) Paris, and an Habilitation for Research Direction from the University of Rennes. From 1985 to Oct. 1997, she has been with FRANCE TELECOM, where she has been involved in various projects in the area of image and video coding and processing for TV, HDTV and multimedia. From Jan. 1990 to mid 1991, she has worked at Bellcore, NJ, USA, as a visiting scientist. Her research interests are signal and image processing, and computer vision.

She has served as Associate Editor for IEEE Trans. on Image Processing (from 2000 to 2003, and from 2014-2016), for IEEE Trans. on Circuits and Systems for Video Technology (from 2004 to 2006), and for IEEE Trans. on Signal Processing (2007-2009). She has served as senior member of the editorial board of the IEEE journal on selected topics in signal processing (2013-2015) and has been senior area editor of IEEE Trans. on Image Processing (2016-2020).

CFA Sensor with $1, 2, 3$ **shots and with** 4 **ISO values:** $100, 400, 800, 1600$**.**			
Block-based ISO	Joint Reconstruction		
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	28.04 ; **28.77** ; 28.72	0.9237 ; **0.9445** ; 0.9441	57.48 ; 61.05 ; **61.41**
Building	**30.09** ; 29.40 ; 29.29	0.8986 ; 0.9340 ; **0.9356**	54.35 ; 56.64 ; **57.32**
Trophy	29.35 ; 31.69 ; **32.59**	0.9301 ; **0.9442** ; 0.9432	**54.05** ; 53.90 ; 52.87
Pendulum	**26.00** ; 25.72 ; 25.51	0.9033 ; **0.9224** ; 0.9222	55.51 ; 59.09 ; **59.72**
Window	30.85 ; 32.27 ; **32.28**	0.8966 ; 0.9397 ; **0.9483**	54.03 ; 57.45 ; **58.53**
Begonia	**30.96** ; 30.10 ; 30.02	0.9359 ; **0.9510** ; 0.9505	56.82 ; **59.99** ; 59.72
Cactus	31.53 ; **33.00** ; 32.87	0.9038 ; 0.9458 ; **0.9505**	54.32 ; 59.13 ; **60.23**
Plane	28.50 ; 31.38 ; **32.61**	0.9147 ; 0.9390 ; **0.9439**	55.12 ; 56.21 ; **56.45**
Row-based ISO	Joint Reconstruction		
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	28.77 ; **30.36** ; 29.88	0.9259 ; **0.9554** ; 0.9533	58.16 ; 61.87 ; **62.25**
Building	**31.23** ; 31.00 ; 30.35	0.8973 ; 0.9296 ; **0.9298**	54.73 ; 57.33 ; **58.21**
Trophy	35.15 ; **36.11** ; 35.56	0.9479 ; **0.9559** ; 0.9517	**55.33** ; 54.92 ; 54.21
Pendulum	28.77 ; **29.58** ; 29.12	0.9082 ; **0.9356** ; 0.9330	55.92 ; 59.47 ; **59.65**
Window	31.25 ; 33.21 ; **33.50**	0.8848 ; 0.9255 ; **0.9314**	54.66 ; 57.58 ; **57.83**
Begonia	32.37 ; 32.76 ; **32.76**	0.9356 ; **0.9516** ; 0.9505	57.04 ; **59.90** ; 59.51
Cactus	32.02 ; 34.25 ; **34.47**	0.9116 ; 0.9476 ; **0.9528**	55.38 ; 59.76 ; **60.78**
Plane	32.80 ; 35.58 ; **36.67**	0.8739 ; 0.9243 ; **0.9528**	55.56 ; 56.57 ; **57.17**
Random ISO	Joint Reconstruction		
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	28.68 ; **30.39** ; 29.87	0.9254 ; **0.9557** ; 0.9540	58.12 ; 62.11 ; **62.17**
Building	**31.24** ; 30.83 ; 30.15	0.8976; 0.9295 ; **0.9299**	54.62 ; 57.50 ; **57.85**
Trophy	31.73 ; **35.98** ; 35.57	0.9197 ; **0.9557** ; 0.9517	54.00 ; **55.11** ; 54.63
Pendulum	28.66 ; **29.67** ; 29.12	0.9090 ; **0.9358** ; 0.9328	55.85 ; 59.64 ; **60.05**
Window	31.23 ; 33.21 ; **33.50**	0.8842 ; 0.9255 ; **0.9316**	54.68 ; 57.59 ; **57.95**
Begonia	32.38 ; **32.79** ; 32.63	0.9358 ; **0.9521** ; 0.9502	56.97 ; 59.66 ; **59.69**
Cactus	32.00 ; 34.28 ; **34.51**	0.9121 ; 0.9480 ; **0.9533**	55.48 ; 59.84 ; **60.85**
Plane	32.77 ; 35.59 ; **36.68**	0.8736 ; 0.9250 ; **0.9528**	55.31 ; 56.24 ; **57.10**
Average	31.09 ; **32.84** ; 32.38	0.9072 ; 0.9409 ; **0.9445**	55.63 ; 58.46 ; **58.79**

Individual ISO reconstruction with 4 shots with ISOs: 100, 400, 800, 1600 with exposure time: 0.004			
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Bike	31.38	0.9324	58.91
Building	31.03	0.9022	54.89
Trophy	32.90	0.9506	54.31
Pendulum	29.88	0.9029	49.05
Window	29.83	0.8613	53.52
Begonia	33.14	0.9467	57.30
Cactus	31.47	0.9030	54.02
Plane	30.37	0.9144	51.40

TABLE VII: Results of the proposed joint reconstruction algorithm in comparison with individual ISO reconstructions for real light fields. The acquisition scheme assumes a CFA sensor, four ISO values ($100, 400, 800$ and 1600) with three different ISO patterns: block-based, Row-based and random. The regularization strength, μ in (9), has been set to 0.0001. These results have been achieved with $1, 2$ and 3 acquisitions (shots). The exposure time has been set to 0.004 seconds. The stopping criterion of the ADMM algorithm is set to 2^{-18}. Moreover, we set $\gamma = 2.2$

APPENDIX

Table VII gives the results per real light field with 1, 2, and 3 shots, for the three ISO patterns (random, block-based and row-based) with ISO values of $100, 400, 800$ and 1600. Note that for the row-based pattern, we only use ISO 100 and 1600. Table VIII includes our results for the synthetic light fields, *Classroom* and *Pabellon*, where we use the row-based pattern but with four ISO values, namely ISO100, ISO400, ISO800, and ISO1600. Finally, Fig. 12 shows the evolution of the reconstruction quality (HDR-VDP2, PU-PSNR, PU-SSIM) with increasing exposure time for 4 light fields and with the random ISO pattern. As shown in this figure, our framework is robust against saturation for most scenes and the quality of the reconstruction starts to fade away when more than 15% of the pixels are saturated, see Fig. 12(d).

CFA Sensor with 3 **shots with ISO values:** $100, 400, 800, 1600$**.**			
Fixed Row-based ISO	Joint Reconstruction		
Light Fields	PU2 PSNR	PU2 SSIM	HDR VDP2
Classroom - baseline 1	25.30	0.9094	57.42
Classroom - baseline 3	27.72	0.9328	56.40
Pabellon - baseline 1	27.12	0.9078	55.51
Pabellon - baseline 3	28.17	0.9272	55.57

TABLE VIII: Results of our proposed joint reconstruction algorithm for the synthetic HDR light field. The acquisition scheme assumes four ISO values ($100, 400, 800, 1600$) varying spatially using Row-based pattern. The exposure time is $t = 0.041$s and 3 shots have been captured. The regularization strength, μ in (9), has been set to 0.0001 for *Classroom* and 0.001 for *Pabellon*. The stopping criterion of the ADMM algorithm has been set to 2^{-16}. Moreover, we set $\gamma = 3.0$.

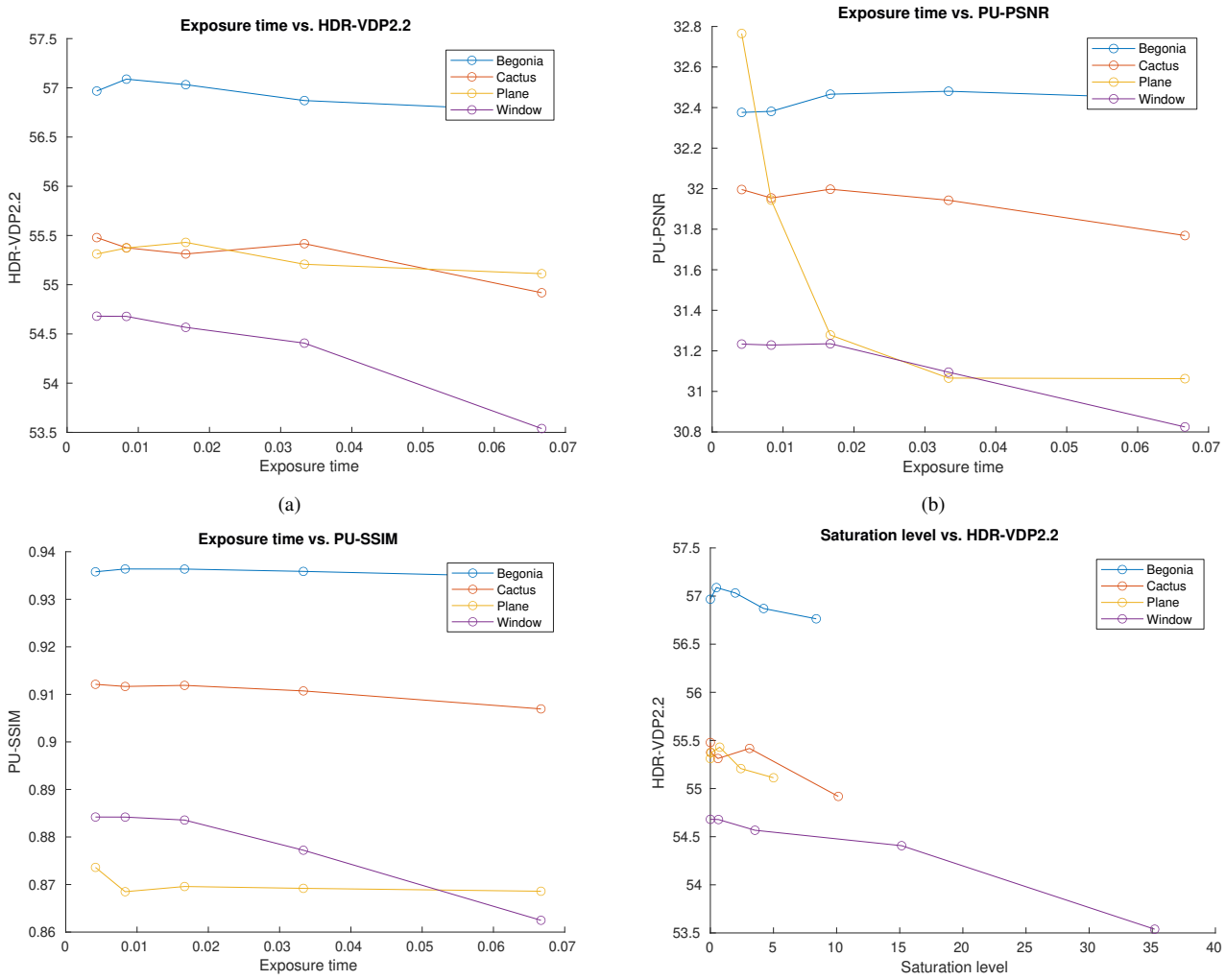


Fig. 12: Evolution of reconstruction quality (HDR-VDP2, PU-PSNR, PU-SSIM) with increasing exposure time for 4 light fields (*Begonia*, *Cactus*, *Plane*, *Window*), with the random ISO pattern. (d) the reconstruction quality as a function of the percentage of saturated pixels (right), with the random ISO patterns.