Time-Multiplexed Coded Aperture Imaging: Learned Coded Aperture and Pixel Exposures for Compressive Imaging Systems

Edwin Vargas[1,*], Julien N.P. Martel[2,*], Gordon Wetzstein[2], Henry Arguello[1]
[1]Universidad Industrial de Santander, Colombia [2]Stanford University, USA
edwin.vargas4@correo.uis.edu.co, [jnmartel,gordonwz]@stanford.edu, henarfu@uis.edu.co

Abstract

Compressive imaging using coded apertures (CA) is a powerful technique that can be used to recover depth, light fields, hyperspectral images and other quantities from a single snapshot. The performance of compressive imaging systems based on CAs mostly depends on two factors: the properties of the mask's attenuation pattern, that we refer to as "codification", and the computational techniques used to recover the quantity of interest from the coded snapshot. In this work, we introduce the idea of using time-varying CAs synchronized with spatially varying pixel shutters. We divide the exposure of a sensor into sub-exposures at the beginning of which the CA mask changes and at which the sensor's pixels are simultaneously and individually switched "on" or "off". This is a practically appealing codification as it does not introduce additional optical components other than the already present CA but uses a change in the pixel shutter that can be easily realized electronically. We show that our proposed time-multiplexed coded aperture (TMCA) can be optimized end to end and induces better coded snapshots enabling superior reconstructions in two different applications: compressive light field imaging and hyperspectral imaging. We demonstrate both in simulation and with real captures (taken with prototypes we built) that this codification outperforms the state-of-the-art compressive imaging systems by a large margin in those applications.

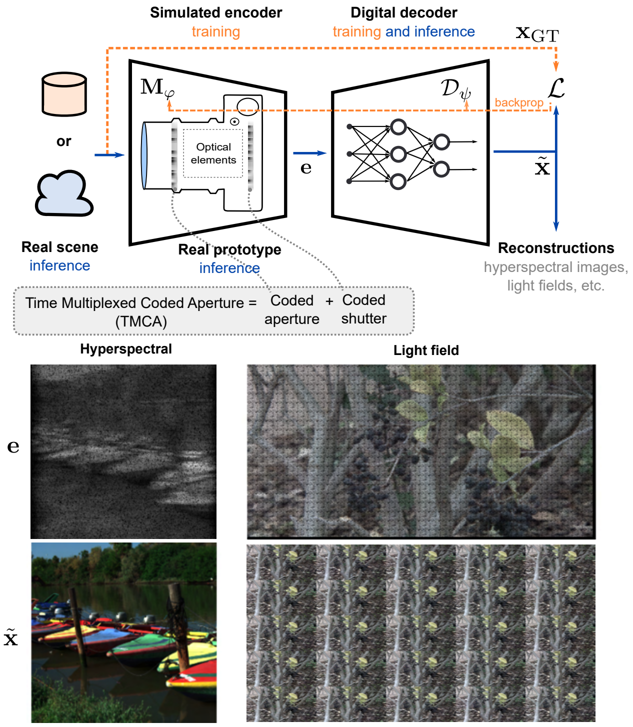

Figure 1. An illustration of the proposed Time Multiplexed Coded Aperture (TMCA) codification in the proposed end-to-end differentiable pipeline. We also show coded snapshot and reconstruction examples for our compressive light field and hyperspectral imaging applications.

1. Introduction

Computational imaging techniques combining the co-design of hardware and algorithms have successfully enabled the development of novel cameras for several applications such as spectral imaging [2], depth imaging [11], light-field imaging [35], or computed tomography [25]. Among those, in particular compressive imaging [13] approaches that aim at multiplexing visual information in a single snapshot have been popularized to reconstruct high dynamic range images, videos, spectral or depth information [36, 58, 12]. They use optical and electronic coding strategies to encode light and use physics-based forward models along with optimization techniques to reconstruct the aforementioned visual quantities. Besides their ability to recover visual information not easily accessible with traditional cameras, another advantage of compressive imaging systems is they can realize sparse measurements, for instance capturing a single coded snapshot to recover a whole light field. This typically yields ill-posed reconstruction problems that can nevertheless be solved by leveraging knowledge about the signal's sparsity and using other priors about the visual quantity to reconstruct. For optimal re-

[*] denotes equal contributions.

construction, compressive sensing theory usually assumes dense uncorrelated measurement matrices. However, this is rarely the case, as those measurement matrices are induced by design constraints and the physical realization of the optical and electronic codifications. As a result, those matrices are almost always sparse and highly structured, thus impeding the performance of compressive imaging systems. Hence, the main lever to improve those systems is to improve their codification.

In particular, many compressive imaging systems employ coded aperture (CA) masks as codification elements [44, 29, 51, 45, 35, 3, 5]. CAs are physical components inserted in the optical system to spatially modulate light intensity. As an example, the coded aperture snapshot spectral imaging system (CASSI) presented in [52, 2] employs a dispersive element behind a CA used to select the spatial locations that get spectrally dispersed. Those CAs can be fabricated using inexpensive technologies (such as chrome-on-quartz) for simple binary codes. However creating multi-valued color codes, yielding much better reconstructions, requires expensive microlithography and coating technologies [3]. Yet another example illustrating the use of CA masks in compressive imaging is in the reconstruction of light fields, for instance by using a CA inserted between the sensor and the objective lens [35]. In this system, correlations in the measurement matrix are created by the similarity between the angular views and typically limit the quality of the light-field reconstructions. These two examples illustrate the need for CA systems that yield better codifications and that are still easy and inexpensive to implement.

This work addresses this challenge by proposing a new codification we call Time-Multiplexed Coded Aperture (TMCA). It induces a better conditioning of those measurement matrices and can be realized using existing components that are widely used in compressive imaging systems. The proposed TMCA consists of using a conventional amplitude CA along with a coded exposure. Coded exposures are temporal modulations of the pixel of a sensor. In a coded exposure, a shutter function turns individual pixels "on" or "off" during the exposure, thus modulating light integration. Pixel-wise shutters can be realized electronically modifying the pixel architecture [34, 53, 32, 50] or using spatial light modulators (SLMs) [20]. More specifically, we generate a TMCA by synchronizing a CA that changes its pattern in time –dubbed as a time-varying CA– and a spatially varying shutter-function realizing a coded exposure. We show that this combination produces a new family of CAs with better codification capabilities that we demonstrate in two specific compressive imaging applications: hyperspectral imaging and light-field imaging.

Furthermore, we propose to learn the TMCA codification, inspired by works in deep optics [34, 11, 36, 57, 40, 55]. We perform the joint optimization of the TMCA codes along with a neural network (NN) used to recover the light fields or hyperspectral images. Our end-to-end (E2E) differentiable approach can be interpreted as an encoder–decoder framework in which the encoder performs the optical-electronic codification and a digital decoder (the NN) is used for reconstruction. The CA and shutter functions are optimized for each application in simulation, considering the specific constraints of the spatial light modulators realizing the CA and of the sensor realizing the coded exposures. After training, the optimized TMCA codes can be deployed to physical devices that can be used to capture real-world scenes.

We summarize the contributions of our work as follows:

- We introduce a new codification for compressive imaging systems called time-multiplexed coded aperture (TMCA).
- We develop new forward models based on the use of TMCAs for two applications: hyperspectral imaging and compressive light field imaging. In the first case, we show TMCA emulates an expensive color filter array, while in the second, the TMCA is an angular sensitive CA resulting in less correlations in the coding of angular views.
- We learn, in simulation, differentiable optical electronic encoders realizing the TMCAs as well as a NN decoders solving the reconstruction problems. We demonstrate that the learned codes compare favorably against baselines that use traditional CA as well as against our own TMCA baselines using non-optimized codes.
- We build two prototypes: a compressive light field imaging system and a hyperspectral imager. We translate the learned TMCA codes to hardware and conduct real experiments showing the better results of TMCA in simulation transposes to real-world systems.

2. Related work

Coded aperture systems employ carefully designed mask patterns to encode incident light. They can be thought of as arrays of pinholes that were developed to improve upon the light efficiency of single pinhole cameras. CA-based systems are widely employed in astronomy or biomedical applications in which lenses cannot easily be fabricated because of the wavelengths at play [14, 15]. Coded snapshots captured through CAs can be decoded using computational techniques to provide sharp, clean images. Recent works have considered CA methods for developing novel image acquisition techniques for depth imaging [29], motion deblurring [44], lensless imaging [59, 5], high-dynamic range [39], video imaging [33]. Furthermore, compressive sensing methods have also been used jointly with CAs: examples include spectral imaging [52, 3], dual-photography [46], light field imaging [35], or image super-resolution [37]. The quality of the reconstruction in all these applications mainly depends on the codification created by the CA. Hence, our work could readily be adapted for any of those applications and we believe, would directly improve them. Here we choose to focus on hyperspectral imaging and light field imaging.

Hyperspectral imaging (HSI) aims at capturing images with a large number of spectral channels (more than the three typical red, green, blue bands) [27]. There are three main approaches for HSI: computed tomography imaging, spectral scanning, and snapshot compressive imaging. Based on a dispersive optical element, such as a prism or a diffraction grating, scanning-based approaches can capture each wavelength of light in isolation through a slit: so-called whiskbroom or pushbroom scanners [8, 43]. While scanning methods yield high spatial and spectral resolution, they are typically slower than other methods. Computed tomography imaging spectrometry [17, 23, 41] was introduced to mitigate this limitation. It employs a diffraction grating that diffracts incident collimated light into patterns in different directions, even though such systems can be real time, this is at the expense of spatial resolution. Finally coded aperture snapshot spectral imaging (CASSI) methods [51, 12, 7] were also introduced for faster captures. Similar to other compressive imaging techniques, those are limited by the codification properties as well the reconstruction algorithms they use. Our work addresses those limitations introducing a new coding strategy and learning the codes in an end-to-end fashion.

Light-field imaging aims at capturing the amount of light passing through every direction in any point in space, practically representing a scene through different "angular perspectives". Early light-field (LF) camera prototypes either used a film sensor using pinholes or microlens arrays [22]. More recently, camera arrays [54, 56, 30, 31] improved the quality of LF captures. However, these approaches may require multiple cameras and snapshots and are often impractical or expensive to build. Using compressive sensing, LF architectures have been proposed with the goal of taking fewer snapshots [4, 6, 35]. Marwah et al. [35] first introduced coded light-field photography (CLFP), in which a light field can be recovered from a single coded measurement obtained with a CA inserted between the sensor and the objective lens. Hirsch et al. [19] proposes another coding solution using angular sensitive pixels (ASP) that has been shown to improve the conditioning of the measurement matrix over CA approaches such as [35]. A limitation of this approach is that it uses specific sensors with ASP of given angular and frequency responses. Our work allows the generation of analogous codifications using a CA, improving the reconstruction but using more flexible codifications (they can be changed) and hardware.

End-to-end optimization and the co-design of optics and algorithms is at the core of computational photography. Automatic differentiation programming tools [42, 1], enable the implementation of fully differentiable pipelines and have fueled this concept with a number of applications such as color imaging and demosaicing [9], extended depth of field imaging [49], depth imaging [11, 18, 57], image classification [10, 38], HDR imaging [36], microscopy [40, 26]. Our work builds on end-to-end optimization techniques, in the line of [34], it does not only optimize the optical encoding but a part of the sensor itself: the coded exposures, together with the NN used for reconstruction.

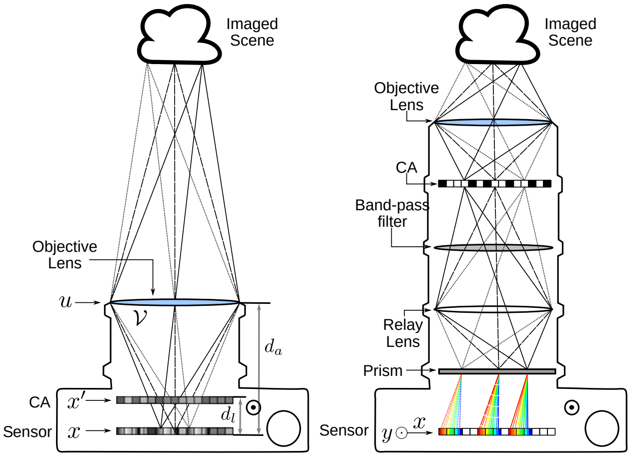

Figure 2. A diagram of the ray optics for our light field imaging (left) and hyperspectral imaging (right) systems using TMCAs.

3. Time-multiplexed coded apertures

The principle of the proposed TMCA consists of synchronizing a time-varying CA with a shutter function realizing a coded exposure. This idea can be used to realize different systems depending on the specificities of the optical setup it is coupled with, i.e, the optical elements in between the CA and the sensor as shown in Figure 1. First, we present the general forward model and codification induced by a TMCA. We then describe how this model can be used for two specific applications using different optical setups: in light-field and hyperspectral imaging, derive the codification for those two systems and demonstrate how the proposed TMCA improves the traditional CAs they use.

3.1. Generalities

We consider the irradiance I, invariant in time, the quantity we reconstruct with our compressive imaging systems. Our proposed TMCA consists of two coding stages.

The first stage optically encodes I into a field g incident to the sensor, using a time-varying CA $T(t)$. We model $g(t)$ as the response of a linear optical system $\mathcal{O}$ given input I:

$$g(t) = \mathcal{O}(T(t), I). \tag{1}$$

As we shall see in section 3.2 and 3.3, the exact form of $\mathcal{O}$ is application dependent but is always a function of the time-varying coded exposure $T(t)$.

The second stage consists of a coded exposure. The irradiance $g(t)$ is captured by the sensor in a single snapshot of exposure time Δt. During this exposure, a spatially-varying shutter function $S_{i,j}(t)$ turns the pixel (i,j) "on" and "off" multiple times, resulting in the coded exposure $e_{i,j}(t)$ in which the integration of $g(t)$ has been modulated:

$$e_{i,j}(t) = \int_{t}^{t+\Delta t} S_{i,j}(t') g_{i,j}(t')\,\mathrm{d}t'. \tag{2}$$

In particular, we consider binary shutter functions $S_{i,j}$ defined on K discrete time slots (sub-exposures of time δt): we have $\Delta t = K \cdot \delta t$. The coded exposure can then be rewritten as

$$e_{i,j} = \sum_{k=0}^{K-1} S_{i,j}^k g_{i,j}^k, \text{with } S_{i,j}^k \in \{0,1\}^k, \tag{3}$$

where $g_{i,j}^k$ and $S_{i,j}^k$ denote the irradiance incident on the sensor and shutter function at pixel (i,j) in the $k-$th time slot. Those slots allow us to synchronize the coded apertures with the coded exposures (both are indexed by k).

Another way to write this discrete model is using the matrix-vector notation $\mathbf{e} = \sum_k \boldsymbol{S}^k \mathbf{g}^k$, where $\mathbf{e}$ and $\mathbf{g}$ are the "vectorized" form of the exposure and coded irradiance and $\boldsymbol{S}$ is the measurement matrix representing the shutter function. The irradiance incident to the sensor is $\mathbf{g}^k = \boldsymbol{O}^k \mathbf{x}$, where the matrix $\boldsymbol{O}^k$ is the point spread function of the application-dependent optical system, also including the coded aperture, and $\mathbf{x}$ represents the irradiance I in its vector form. Using this notation, the forward model of the proposed TMCA can be simply written as

$$\mathbf{e} = \sum_{k=0}^{K-1} \boldsymbol{S}^k \boldsymbol{O}^k \mathbf{x} = \mathbf{M}\mathbf{x}, \tag{4}$$

defining the overall measurement matrix of our compressive imaging system $\mathbf{M} = \sum_{k=0}^{K-1} \boldsymbol{S}^k \boldsymbol{O}^k$. Recovering $\mathbf{x}$ given the coded snapshot $\mathbf{e}$ amounts to solving an inverse problem.

In the following sections, we use this same general model for two different applications. In our first application of compressive light field imaging, we aim at recovering light fields and the irradiance I (and thus $\mathbf{x}$) considers multiple view angles. Our second application targets hyperspectral imaging. In that case, I (and $\mathbf{x}$) considers multiple frequency bands.

3.2. TMCA for compressive light field imaging

We consider the same optical setup proposed for compressive light field imaging by Marwah et al. [35]. A coded aperture mask T is placed between the objective lens and the sensor, at a distance d_l from the latter. The incident field $g(x,t)$ at the sensor is the spatially modulated light field projected along its angular dimension taken over the aperture area $\mathcal{V}$:

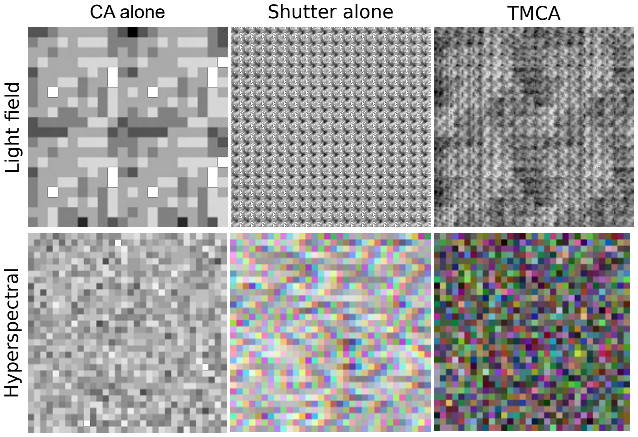

Figure 3. Illustrations of the codifications for $K = 8$. Right: full codification as in Eqs.(13) and (7). Left and middle: codifications when $\boldsymbol{S} = \mathrm{Id.}$ (left) and $\boldsymbol{T} = \mathrm{Id.}$ (middle). Even though the codes are binary, $K = 8$ yields 2^8 possible shutter / CA values.

$$g(x) = \int_{\mathcal{V}} l(x,u)\, T\big(x + s(u-x)\big)\,\mathrm{d}u, \tag{5}$$

where $s = d_l/d_a$, with d_a the distance from the sensor to the aperture plane, is the shear of the mask pattern with respect to the incident light field $l(x,u)$. Similar to [30], we adopt a two-plane parameterization for the light field where x is the 2D spatial dimension on the sensor plane and u denotes the 2D position on the aperture plane (see Fig. 2).

Now considering the time-varying coded aperture T and the shutter function S, introduced in Section 3.1, the measurement model yielding the exposure e is

$$e(x) = \int_{\mathcal{V}} \int_{\Delta t} S(x,t')\, l(x,u)\, T(x + s(u-x), t')\,\mathrm{d}u\,\mathrm{d}t'. \tag{6}$$

By defining the TMCA $\hat{T}$ as

$$\hat{T}(x,u) = \int_{\Delta t} S(x,t')\, T(x + s(u-x), t')\,\mathrm{d}t', \tag{7}$$

the model in Equation (6) can be simply rewritten as

$$e(x) = \int_{\mathcal{V}} l(x,u)\, \hat{T}(x,u)\,\mathrm{d}u. \tag{8}$$

The model from [35] described in Equation (5) does not include the coded exposure. Comparing the latter with our TMCA model in Equation (7), we note that our model can be seen as an equivalent coded aperture we denoted $\hat{T}$.

Furthermore, using a change of variable, we can show that the proposed TMCA for compressive light field imaging can be expressed in the coded aperture plane as

$$\hat{T}(x',u) = \int_{\Delta t} S(x' + \hat{s}(u-x'), t') T(x',t')\,\mathrm{d}t', \tag{9}$$

using the spatial coordinates x' in the coded aperture plane (see Fig.2) and with $\hat{s} = d_\ell/(d_a - d_\ell)$.

Interestingly, this shows that using the proposed TMCA, the equivalent CA's pixels can be seen as responding differently to rays coming from different angles. If the shutter function is removed and the CA remains constant in time, then the TMCA reduces to [35], where all the pixels respond equally for all angles. If we were now to consider the shutter function alone, without the CA, the time-modulation in the sensor plane averages all the views, which brings no clear coding advantage if the scene is static in time (since all the views would be the same). In the top row of Figure 3 we plot the codification in the coded aperture plane.

3.3. TMCA for compressive hyperspectral imaging

We use an optical design similar to the coded aperture spectral snapshot imager (CASSI) proposed in [52] for hyperspectral compressive imaging. In this architecture, spectral dispersion is achieved using a prism between the lens and the sensor. The quantity we aim at recovering is the irradiance $I(x, y, \lambda)$. Note that we are now explicitly considering a second spatial dimension y (and will assume the prism disperses in the x dimension) and the spectral dimension λ. Similarly, we denote the spatial dependency of the coded aperture $T(x, y)$. The field impinging the sensor is now also dependent on the optical response of the prism h as well the spectral response of the sensor κ:

$$g(x,y) = \iiint T(x',y')\, I(x',y',\lambda)\; h(x - \mathfrak{s}(\lambda) - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda, \quad (10)$$

where $\mathfrak{s}(\lambda)$ is the wavelength dependent spatial shift induced by the prism.

Using the shutter function S to create the TMCA, the optically encoded field g yields the coded exposure

$$e(x,y) = \int_{\Delta t} S(x,y,t') \iiint T(x',y',t')\, I(x',y',\lambda)\; h(x - \mathfrak{s}(\lambda) - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda\, \mathrm{d}t'. \quad (11)$$

Since h is the propagation through unit magnification imaging optics and a dispersive element with linear dispersion, the impulse response can be expressed as $h(x - \mathfrak{s}(\lambda) - x', y - y') = \delta(x - \lambda - x', y - y')$. After substituting this expression in (11) simplifying and rearranging (see supplemental), we can express the coded measurements as

$$e(x,y) = \iiint \hat{T}(x',y',\lambda)\, I(x',y',\lambda)\; \delta(x - \lambda - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda, \quad (12)$$

that use the TMCA $\hat{T}$ defined as

$$\hat{T}(x',y',\lambda) = \int_{\Delta t} S(x' + \lambda, y', t') T(x',y',t')\, \mathrm{d}t'. \quad (13)$$

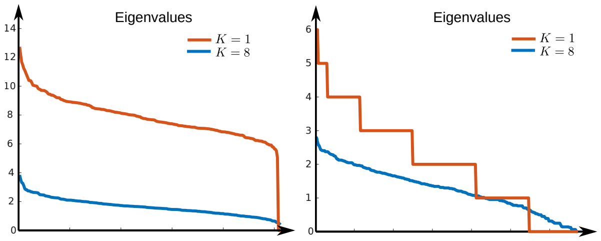

Figure 4. Eigenvalues distribution sorted in descending order for the compressive light field (left) and compressive spectral imaging system (right) for the case of $K = 1$ (traditional, no coded exposure) compared with the case of $K = 8$ (more integration slots in the TMCA).

In summary, adding the coded exposure to the model of [52] generates a TMCA with a new coded aperture $\hat{T}$. The dependency of Equation (13) in λ shows the proposed TMCA emulates a color coded aperture which would otherwise be expensive to create. Importantly, note this is only true when both the shutter function and CA are jointly employed. If the shutter function is a constant, this is exactly the model in [52]: there is no spectral response of the CA. On the other hand, if the CA is removed, $\hat{T}$ exhibits a spectral response that shares the same code for every wavelength that is simply shifted in the x direction depending on the wavelength.

A final advantage of the TMCA codification is that even in the case we would restrict the coded exposure or coded aperture to binary values, for instance because those would be simpler to realize physically, the proposed codification can still produce spatio-spectral patterns with non-binary values of attenuation. The bottom row of Fig. 3 depicts the codification of the proposed TMCA.

3.4. Conditioning of the measurement matrices

We empirically analyze the conditioning of the proposed TMCA codifications. To do so, we consider discretized versions of Equations (12) and (8) (derived in the supplemental). Both reduce to the general form presented in Equation (4), where $\mathbf{M}$ is the measurement matrix. The eigenvalue distribution of this matrix informs us about its conditioning and thus our ability to solve the inverse problem, that is recovering $\mathbf{x}$ from $\mathbf{e}$.

Figure 4 shows the two TMCAs for different number of discrete slots K in the shutter function (Section 3.1). The cases shown for $K = 1$ are equivalent to using no shutter function (a single slot means a single integration). Those are the measurement matrices induced by the traditional compressive light field in [35] and compressive spectral imaging in [52]. The plots for both applications show a better conditioning of the measurement matrices for $K = 8$ (TMCA) compared with $K = 1$ (traditional CA). The ratio between the lowest and highest eigenvalues is lower for $K = 8$ and the distribution is more uniform (the eigenvalues decay less rapidly), indicating TMCA is a better codification.

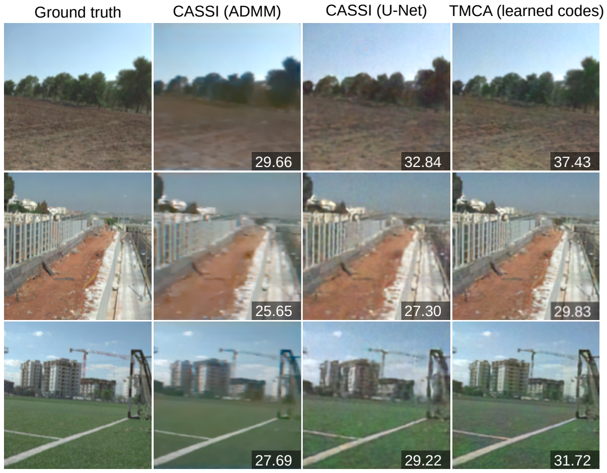

Figure 5. Examples of compressive spectral imaging reconstructions in simulation. The PSNR (dB) between the reconstructions and the ground truth images is shown in the lower-right corner.

Methods	PSNR(↑)	UIQI(↑)	SAM(↓)	ERGAS(↓)	DD(↓)
CASSI (ADMM)	27.40	0.938	22.42	9.56	0.031
CASSI (U-Net)	29.66	0.968	15.99	7.04	0.022
CASSI (E2E learned)	30.23	0.971	15.11	6.72	0.020
TMCA (random)	31.39	0.978	13.01	5.74	0.019
TMCA (learned)	**32.72**	**0.981**	**11.92**	**5.27**	**0.016**

Table 1. Compressive spectral imaging: Comparison of the proposed TMCA with baselines on the ICVL 1 dataset.

4. End-to-end optimization: learning codes and reconstructions jointly

We consider an end-to-end approach in which we optimize both the optical electronic encoder realizing the TMCA with a NN decoder solving the inverse problem (Figure 1). Using NNs to implement our differentiable decoders presents three main advantages: 1) they are differentiable: the error is propagated back to the codes and can jointly optimize both the encoder and decoder, 2) they embed priors that are task and dataset dependent, 3) they use a single feed-forward pass which is, in many cases, faster than traditional iterative methods (such as dictionary based methods).

Encoders There are two main challenges to address in the end-to-end optimization of our TMCA. First, since both the CA and shutter functions are implemented in hardware (see Section 5), they are subject to the real devices' and systems' constraints. In particular our CAs and shutter functions must be binary valued. A second challenge is that we discretized the time domain in time slots to easily synchronize them. This essentially means the parameters representing those functions need to encode the slot at which sub-exposure start and stop. Without any further constraint on the functional form of the CA or coded exposure, the number of those parameters grows as the number of slots increases, preventing us from using too many slots.

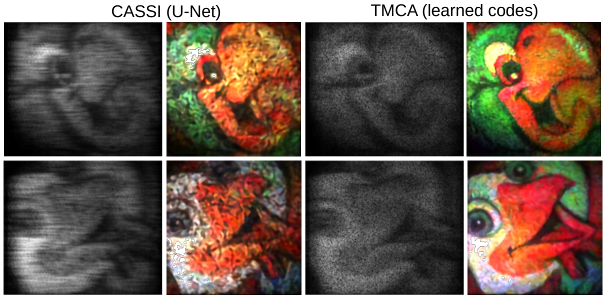

Figure 6. Examples of real captures of coded snapshots for hyperspectral imaging and their reconstructions.

To optimize the TMCA under those constraints we use an approach similar to the one presented in [34]. A forward pass through the encoder implements the exact discrete, binary valued model but the backward pass used to optimize its parameters is mismatched and implemented considering a continuous approximation of the discrete forward model. As an example, hard thresholds used for quantization in the forward pass are considered to be sigmoid functions in the backward pass. This "forward-backward mismatch" approach enables gradients to flow backward to update the encoder's parameters while keeping an exact forward model that can be directly translated into hardware.

In the following, we denote our encoders as $\mathbf{M}_\phi$, where ϕ are the set of parameters that encode the discrete forward model. In TMCA, those parameters are essentially K arrays of $M \times N$ pixels representing the time varying CA (on K slots), and the K arrays of $M' \times N'$ sensor's pixel representing the shutter functions.

Decoders Many types of differentiable decoders have been considered for end-to-end optimization [49, 57, 40]. For the hyperspectral imaging application we use a vanilla U-Net preceded by a lifting of the measurement using a back-projection with the transposed of the measurement matrix. The measurement $\mathbf{e}$ generated by our hyperspectral encoder is a single 2D snapshot of size $M \times (N + L - 1)$ while the hyperspectral image $\mathbf{x}$ we wish to reconstruct is a cube of size $M \times N \times L$. U-Nets operate by translating one domain to another of same dimension. Therefore, we first lift the measurements $\mathbf{e}$ in the spectral domain using the transpose operator $\mathbf{M}^T$ creating $\mathbf{e}' = \mathbf{M}^T\mathbf{e}$, which is then fed to the U-Net decoder to reconstruct the hyperspectral image $\tilde{\mathbf{x}}$. For the compressive light field application we used the unrolled optimization network proposed by [16]. We denote the function produced by those NNs $\mathcal{D}_\psi$ where ψ are their learnable parameters.

At training time, a ground-truth measurement $\mathbf{x}_{\text{GT}}$ is sampled from a dataset of N hyperspectral images or light fields. It is then encoded via the forward model as $\mathbf{e} = \mathbf{M}_\varphi \mathbf{x}_{\text{GT}}$. The decoder proposes a reconstruction $\tilde{\mathbf{x}} = \mathcal{D}_\psi(\mathbf{e})$. To learn the parameters $\{\varphi, \psi\}$ of our encoder–decoder architecture we optimize the loss function $\mathcal{L}$ to

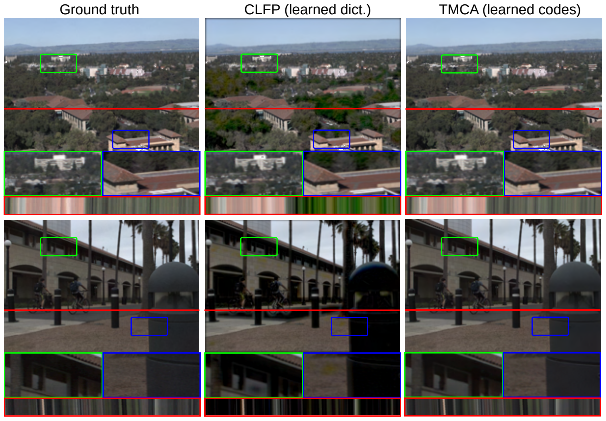

Figure 7. Examples of light-field reconstructions (central view) in simulation (shown with zoom-ins) comparing CLFP [35] and our proposed TMCA with ground truth.

minimize the discrepancy between the reconstructed signals $\tilde{\mathbf{x}} = \mathcal{D}_\psi \circ \mathbf{M}_\varphi(\mathbf{x}^n_{\text{GT}})$ and the ground-truth $\mathbf{x}_{\text{GT}}$:

$$\underset{\phi,\psi}{\text{argmin}} \sum_{n=1}^{N} \mathcal{L}\big(\mathbf{x}^n_{\text{GT}}, \mathcal{D}_\psi \circ \mathbf{M}_\varphi(\mathbf{x}^n_{\text{GT}})\big). \tag{14}$$

where $\circ$ stands for the composition of functions. In practice, we choose $\mathcal{L}$ to be a L_2 or L_1 norm, but this could also be chosen to be a high-level perceptual loss function such as a VGG loss [48] or analogous.

At inference in simulation, the pipeline is run with $\mathbf{x}_{\text{GT}}$ sampled from a test set. For real captures, the learnt parameters of the encoder are directly translated from simulation to hardware. This is possible because the encoders are, by construction, designed to emulate physically realizable CAs and shutter functions.

5. Results

We show results of the proposed TMCA for the two applications we consider in this work: compressive light-field and hyperspectral imaging. We demonstrate both results in simulations and on real captures taken with two prototype systems we built. All our models are implemented in Pytorch [42] and trained on Titan X GPU using the ADAM optimizer [28], complete training details and photographs of our setups are presented in the supplemental.

5.1. Coded hyperspectral Imaging

Simulations We learn the model for hyperspectral imaging using the ICVL dataset. It consists of 200 spectral images. We randomly select 160 hyperspectral images for training, 20 for validation, and 20 for testing, cropped at a size of 256 × 256 with $L = 12$ spectral bands. We set the number of time slots in the TMCA encoder to $K = 8$. The U-Net is trained for 500 epochs.

We compare the proposed TMCA codification against four different baselines: a) the traditional CASSI codification using random binary patterns and reconstructed using the alternative direction method of multipliers (ADMM) b) the CASSI codification using a trained U-Net as a decoder c) the proposed TMCA codification and reconstruction pipeline using random (non-optimized) codes d) the CASSI codification with joint learned codification and trained U-Net as a decoder and e) our full TMCA codification with learned codes. The quantitative results performed on our random test fold of ICVL are presented in Table 1, they show that on all the metrics we evaluated on (PSNR, UIQI, SAM, ERGAS, and DD, see supplemental for details) our full TMCA pipeline performs better than all the other baselines. We show qualitative results of a few reconstructed hyperspectral images in Figure 5. Those illustrate how the proposed TMCA appears closer to the ground truth both in terms of spatial accuracy (they are less blurry) as well as spectral accuracy (colors match better).

Methods	PSNR(↑)	SSIM(↑)
CLFP[35] (learned dict.)	30.06	0.82
CLFP[35] (deep net.[16])	32.43	0.91
CLFP[35] (E2E learned.)	33.21	0.91
TMCA (random codes)	34.03	0.93
TMCA (opt. codes)	**34.89**	**0.94**

Table 2. Compressive light field imaging: comparison of the proposed TMCA against baselines using the aggregated Lytro dataset.

Real captures We built a prototype of our system to evaluate the proposed TMCA approach for real-world scenes. The system consists of an achromatic objective lens with 50mm focal length (Thorlabs AC254-050-A-ML), a digital micromirror device (DMD DLi4120), an F/8 relay lens, a custom double Amici prism with center wavelength 550nm captured through a monochrome CCD sensor Stingray F-080B with 4.65μm pixel size. The same prototype was used to capture measurements for the traditional CASSI codification and the proposed TMCA. The spectral images are reconstructed using the U-Nets trained in simulation, but using the real measurements. Results showing the capture of a book cover are shown in Figure 6 demonstrating that TMCA imaging can be implemented in a real-world hyperspectral imaging prototype that still presents a higher visual quality than the traditional CASSI.

5.2. Compressive Light Field Imaging

Simulations We learn our end-to-end TMCA for compressive light field imaging on a dataset aggregating real-world and synthetic light fields (LF). We use 100 real captured LF of size 7 × 7 × 376 × 541 from the Kalantari et al. Lytro dataset [24], 22 synthetic LF images of size 5 × 5 × 512 × 512 from [47], and 33 synthetic LF images of size 5 × 5 × 512 × 512 from [21]. We randomly split this aggregated dataset in 110 LF images for training, 20 for validation, and 25 for testing.

In our experiment, we reconstruct 5×5 angular views from a single snapshot with a resolution of 480×270 pixels. We use randomly cropped patches of those images of spatial size 11×11 and consider 5×5 angular views for training: randomly cropping the 4D LF into patches increases the number of samples at training while reducing the memory requirements to process an entire light field. The decoder is the deep spatial-angular convolutional sub-network proposed in [16], it is trained for 500 epochs. After training, we reconstruct overlapping 4D patches that are then merged with a median filter.

We compare the proposed TMCA codification against four different baselines similar to the compressive imaging application: a) a reconstruction using the traditional dictionary learning and reconstruction approach of [35], b) the same codification but using the deep network decoder from [16] c) our TMCA with random (non-optimized codes) using the deepnet decoder of [16] d) E2E optimization of [35] with decoder [16] e) our full TMCA codification pipeline with learned codes. The results are compiled in Table 2, showing that on the two metrics we evaluated on (PSNR and SSIM) the TMCA is superior to the other baselines. Qualitative results are shown in Figure 7 where two recovered LFs of the testing set (Kalantari Lytro LFs) are reconstructed and compared with the ground truth.

Real captures We assess the proposed TMCA approach in a real experiment. We use a liquid crystal on silicon (LCoS) display (HOLOEYE PLUTO-2.1 LCoS SLM) where each pixel can independently change the polarization state of the incoming light field, in conjunction with a polarizing beam splitter and relay optics. As a single pixel on the LCoS cannot be well resolved with our setup, we treat blocks of 4×4 LCoS pixels as macro pixels, resulting in a CA of 480×270. We reimage the LCoS with an SLR camera lens (Canon EF-S $18-55$ f/$4-5.6$ IS STM) which is not focused on the LCoS but in front of it, thereby optically placing the (virtual) image sensor behind the LCoS plane. A Canon EF 80 mm, f/5.6 II lens is used as the imaging lens and focused at a distance of 60cm. We follow the same procedure as in [35] to adjust the distance between the mask (LCoS plane) and the virtual image sensor for capturing light fields with 5×5 angular resolution.

The real captured measurements are reconstructed using the same deep networks as in simulation. The measurements and reconstructions are shown in Figure 8 which shows that the TMCA enables reconstructions featuring better spatial resolution in the angular views.

6. Discussion

We introduced a new coding strategy dubbed Time-Multiplexed Coded Aperture (TMCA) and demonstrated it improves CA codifications in two applications: compressive hyperspectral imaging and light-field imaging. TMCA improves upon traditional CA without introducing additional optical elements. It adds a coded shutter synchronized with the CA. The coded shutter can be simply realized electronically, using dedicated sensors or bursts of multiple snapshots averaged together. We optimized TMCA codes in an end-to-end optimization approach and demonstrated in both synthetic and real experiments that the proposed TMCA yields largely superior reconstruction quality compared with traditional CA approaches. While a limitation of CA systems are their low light efficiency, which could be detrimental in low-light scenarios or high-speed imaging, we believe our system could be further engineered to be real-time (our captures for $K = 8$ take 8×20ms and reconstructions take about a second), thus enabling high-quality videos in those applications. We believe the proposed TMCA would naturally extend to many other applications that make use of CAs and improve their reconstruction quality.

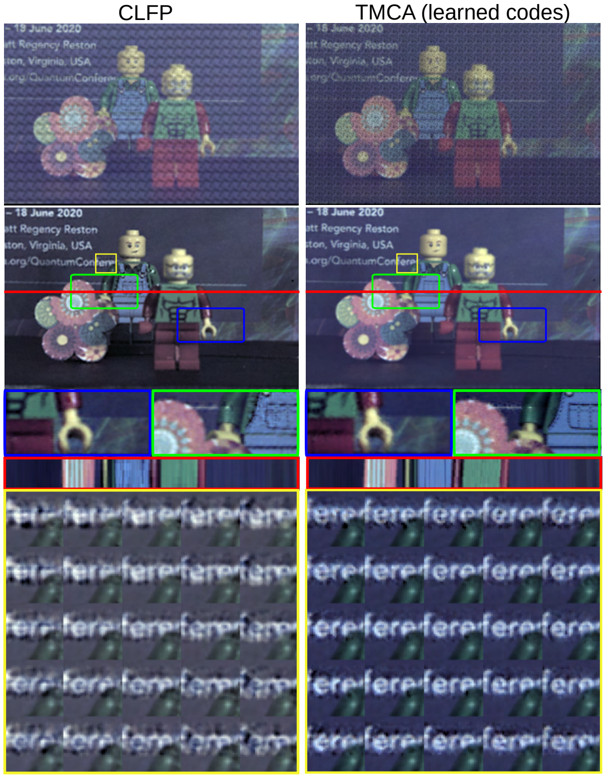

Figure 8. Real captures of coded snapshots for light fields comparing CLFP [35] and the proposed TMCA with their reconstructions.

Acknowledgments

J.N.P.M was supported by a Swiss National Foundation (SNF) Fellowship (P2EZP2_181817), G.W. was supported by an NSF Award (1839974), and a PECASE by the ARL. H.A. was supported by the Fullbright 2019 Visiting Scholar Program.

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016. 3

[2] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2013. 1, 2

[3] Henry Arguello and Gonzalo R Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908, 2014. 2

[4] Amit Ashok and Mark A Neifeld. Compressive light field imaging. In *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, volume 7690, page 76900Q. International Society for Optics and Photonics, 2010. 3

[5] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3):384–397, 2016. 2

[6] S Derin Babacan, Reto Ansorge, Martin Luessi, Pablo Ruiz Matarán, Rafael Molina, and Aggelos K Katsaggelos. Compressive light field sensing. *IEEE Transactions on image processing*, 21(12):4746–4757, 2012. 3

[7] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyperspectral imaging using a prism. *ACM Transactions on Graphics (TOG)*, 36(6):1–12, 2017. 3

[8] Nicola Brusco, S Capeleto, M Fedel, Anna Paviotti, Luca Poletto, Guido Maria Cortelazzo, and G Tondello. A system for 3d modeling frescoed historical buildings with multispectral texture information. *Machine Vision and Applications*, 17(6):373–393, 2006. 3

[9] Ayan Chakrabarti. Learning sensor multiplexing design through back-propagation. In *Advances in Neural Information Processing Systems*, pages 3081–3089, 2016. 3

[10] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific reports*, 8(1):1–10, 2018. 3

[11] Julie Chang and Gordon Wetzstein. Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10193–10202, 2019. 1, 2, 3

[12] Claudia V. Correa, Henry Arguello, and Gonzalo R. Arce. Snapshot colored compressive spectral imager. *J. Opt. Soc. Am. A*, 32(10):1754–1763, Oct 2015. 1, 3

[13] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006. 1

[14] Edward E Fenimore and Thomas M Cannon. Coded aperture imaging with uniformly redundant arrays. *Applied optics*, 17(3):337–347, 1978. 2

[15] Stephen R Gottesman and E Edward Fenimore. New family of binary arrays for coded aperture imaging. *Applied optics*, 28(20):4344–4352, 1989. 2

[16] Mantang Guo, Junhui Hou, Jing Jin, Jie Chen, and Lap-Pui Chau. Deep spatial-angular regularization for compressive light field reconstruction over coded apertures. In *European Conference on Computer Vision*, pages 278–294. Springer, 2020. 6, 7, 8

[17] Ralf Habel, Michael Kudenov, and Michael Wimmer. Practical spectral photography. In *Computer graphics forum*, volume 31, pages 449–458. Wiley Online Library, 2012. 3

[18] Harel Haim, Shay Elmalem, Raja Giryes, Alex M Bronstein, and Emanuel Marom. Depth estimation from a single image using deep learned phase coded mask. *IEEE Transactions on Computational Imaging*, 4(3):298–310, 2018. 3

[19] Matthew Hirsch, Sriram Sivaramakrishnan, Suren Jayasuriya, Albert Wang, Alyosha Molnar, Ramesh Raskar, and Gordon Wetzstein. A switchable light field camera architecture with angle sensitive pixels and dictionary-based sparse coding. In *2014 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2014. 3

[20] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *2011 International Conference on Computer Vision*, pages 287–294. IEEE, 2011. 2

[21] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pages 19–34. Springer, 2016. 7

[22] Frederic E Ives. Parallax stereogram and process of making same., Apr. 14 1903. US Patent 725,567. 3

[23] William R Johnson, Daniel W Wilson, Wolfgang Fink, Mark S Humayun, and Gregory H Bearman. Snapshot hyperspectral imaging in ophthalmology. *Journal of biomedical optics*, 12(1):014036, 2007. 3

[24] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–10, 2016. 7

[25] Willi A Kalender. X-ray computed tomography. *Physics in Medicine & Biology*, 51(13):R29, 2006. 1

[26] Michael Kellman, Emrah Bostan, Michael Chen, and Laura Waller. Data-driven design for fourier ptychographic microscopy. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2019. 3

[27] Min H Kim. 3d graphics techniques for capturing and inspecting hyperspectral appearance. In *2013 International Symposium on Ubiquitous Virtual Reality*, pages 15–18. IEEE, 2013. 3

[28] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[29] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70–es, 2007. 2

[30] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996. 3, 4

[31] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM SIGGRAPH 2008 papers*, pages 1–10. 2008. 3

[32] Yi Luo, Jacky Jiang, Mengye Cai, and Shahriar Mirabbasi. Cmos computational camera with a two-tap coded exposure image sensor for single-shot spatial-temporal compressive sensing. *Optics express*, 27(22):31475–31489, 2019. 2

[33] Roummel F Marcia, Zachary T Harmany, and Rebecca M Willett. Compressive coded aperture imaging. In *Computational Imaging VII*, volume 7246, page 72460G. International Society for Optics and Photonics, 2009. 2

[34] Julien NP Martel, Lorenz Mueller, Stephen J Carey, Piotr Dudek, and Gordon Wetzstein. Neural sensors: Learning pixel exposures for hdr imaging and video compressive sensing with programmable sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2, 3, 6

[35] Kshitij Marwah, Gordon Wetzstein, Yosuke Bando, and Ramesh Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Transactions on Graphics (TOG)*, 32(4):1–12, 2013. 1, 2, 3, 4, 5, 7, 8

[36] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. *2020 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'20)*, 2019. 1, 2, 3

[37] Ankit Mohan, Xiang Huang, Jack Tumblin, and Ramesh Raskar. Sensing increased image resolution using aperture masks. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 2

[38] Alex Muthumbi, Amey Chaware, Kanghyun Kim, Kevin C Zhou, Pavan Chandra Konda, Richard Chen, Benjamin Judkewitz, Andreas Erdmann, Barbara Kappes, and Roarke Horstmeyer. Learned sensing: jointly optimized microscope hardware for accurate image classification. *Biomedical optics express*, 10(12):6351–6369, 2019. 3

[39] Shree K Nayar and Vlad Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *null*, page 1168. IEEE, 2003. 2

[40] Elias Nehme, Daniel Freedman, Racheli Gordon, Boris Ferdman, Lucien E Weiss, Onit Alalouf, Tal Naor, Reut Orange, Tomer Michaeli, and Yoav Shechtman. Deepstorm3d: dense 3d localization microscopy and psf design by deep learning. *Nature Methods*, 17(7):734–740, 2020. 2, 3, 6

[41] Takayuki Okamoto, Akinori Takahashi, and Ichirou Yamaguchi. Simultaneous acquisition of spectral and spatial intensity distribution. *Applied Spectroscopy*, 47(8):1198–1202, 1993. 3

[42] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 3, 7

[43] Wallace M Porter and Harry T Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Imaging Spectroscopy II*, volume 834, pages 22–31. International Society for Optics and Photonics, 1987. 3

[44] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM SIGGRAPH 2006 Papers*, pages 795–804. 2006. 2

[45] Aswin C Sankaranarayanan, Pavan K Turaga, Richard G Baraniuk, and Rama Chellappa. Compressive acquisition of dynamic scenes. In *European Conference on Computer Vision*, pages 129–142. Springer, 2010. 2

[46] Pradeep Sen and Soheil Darabi. Compressive dual photography. In *Computer Graphics Forum*, volume 28, pages 609–618. Wiley Online Library, 2009. 2

[47] Jinglei Shi, Xiaoran Jiang, and Christine Guillemot. A framework for learning depth from a flexible subset of dense and sparse light field views. *IEEE Transactions on Image Processing*, 28(12):5867–5880, 2019. 7

[48] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 7

[49] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 3, 6

[50] Toshiki Sonoda, Hajime Nagahara, Kenta Endo, Yukinobu Sugiyama, and Rin-ichiro Taniguchi. High-speed imaging using cmos image sensor with quasi pixel-wise exposure. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2016. 2

[51] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008. 2, 3

[52] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008. 2, 5

[53] Mian Wei, Navid Sarhangnejad, Zhengfan Xia, Nikita Gusev, Nikola Katic, Roman Genov, and Kiriakos N Kutulakos. Coded two-bucket cameras for computer vision. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 54–71, 2018. 2

[54] Gordon Wetzstein, Ivo Ihrke, and Wolfgang Heidrich. On plenoptic multiplexing and reconstruction. *International journal of computer vision*, 101(2):384–400, 2013. 3

[55] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David AB Miller, and Demetri Psaltis. Inference in arti-

ficial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47, 2020. 2

[56] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. In *ACM SIGGRAPH 2005 Papers*, pages 765–776. 2005. 3

[57] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2019. 2, 3, 6

[58] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021. 1

[59] Assaf Zomet and Shree K Nayar. Lensless imaging with a controllable aperture. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 339–346. IEEE, 2006. 2