Diffraction-Unlimited Imaging Based on Conventional Optical Devices

Nicolas Ducros, Aurélien Bourquard

Diffraction-unlimited imaging based on conventional optical devices

NICOLAS DUCROS[1,4] **AND AURÉLIEN BOURQUARD**[2,3,5]

[1]*Univ Lyon, INSA Lyon, UCBL, CNRS 5220 INSERM U1206, CREATIS, France*
[2]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, USA*
[3]*Biomedical Image Technologies, Universidad Politécnica de Madrid and CIBER-BBN, Madrid, Spain*
[4]*Corresponding author: nicolas.ducros@creatis.insa-lyon.fr*
[5]*Corresponding author: aurelien@mit.edu*

Abstract: We propose a computational paradigm where off-the-shelf optical devices can be used to image objects in a scene well beyond their native optical resolution. By design, our approach is generic, does not require active illumination, and is applicable to several types of optical devices. It only requires the placement of a spatial light modulator some distance from the optical system. In this paper, we first introduce the acquisition strategy together with the reconstruction framework. We then conduct practical experiments with a webcam that confirm that this approach can image objects with substantially enhanced spatial resolution compared to the performance of the native optical device. We finally discuss potential applications, current limitations, and future research directions.

1. Introduction

The resolution of an imaging system, *i.e.*, its ability to separate points that are located at small angular positions, is limited by the density of the sensors and by diffraction. Following advances in signal processing and computer science, computational approaches have been proposed to address cases where the number of sensors is the limiting factor. In particular, the compressed-sensing framework allows images to be reconstructed from substantially fewer sensors than defined by the Shannon-Nyquist sampling limit [1, 2]. A practical implementation of compressed sensing is seen in the single-pixel camera, where a scene is imaged based on a single detector [3]. Applications of single-pixel imaging include microscopy [4], terahertz imaging [5], fluorescence lifetime imaging [6], time-resolved hyperspectral imaging [7], Raman imaging [8], and phase imaging [9]—see [10] for a recent review. In parallel to this trend, however, the spatial density of sensors has increased substantially, with native camera resolutions in standard mobile phones now exceeding ten megapixels. In most consumer and professional optical devices, computational methods that compensate for a lack of sensors are not critical.

The resolution of an imaging device is not only limited by the density of the sensors, but also by the diffraction that occurs inside the optical system. This physical limit has been extensively addressed in fluorescence microscopy. With the advent of super-resolution techniques, different technologies are now available, each of which comes with several variants [11]. These methods rely on complex illumination schemes (*e.g.*, stimulated emission depletion [12]) or photocontrol of the sample (*e.g.*, photoactivated localization microscopy [13], stochastic optical reconstruction microscopy [14]), and they are tailored to fluorescence imaging. More generic resolution-enhancement techniques have also been developed for non-fluorescent objects (*e.g.*, digital holography [15], synthetic-aperture microscopy [16], Fourier ptychography [17]). However, when the object is too far away to be accessed (*e.g.*, in remote sensing or astronomy), none of these approaches can be applied. In this case, only post-processing approaches can be implemented, to deconvolve the instrument response [18]. While deconvolution can improve the image quality when strong prior information is available (*e.g.*, point-like objects), it leads to relatively poor results for extended objects of unknown structure [19].

In this paper, we propose a computational imaging approach for objects that are too far away

to be illuminated or accessed, which allows them to be imaged beyond the limit of diffraction. Our concept is highly flexible, in the sense that it can be applied to any conventional off-the-shelf optical device with minimal hardware modifications and with the addition of a spatial light modulator (SLM) some distance from the optical device. After acquisition of a sequence of images for different SLM patterns, the object can be reconstructed through a simple procedure. To the best of our knowledge, this is the first time that a practical set-up based on these techniques has been shown to break the diffraction limit of an off-the-shelf optical device.

In Section 2, we specify the nature of the imaging problem under consideration. In Section 3, we introduce the joint acquisition and reconstruction paradigm, discuss its intrinsic invariance to diffraction and sampling, and describe an effective implementation. In Section 4, we demonstrate the relevance of the proposed approach through practical optical implementation that is based on a webcam, with a comparison of the results with respect to the conventional case where no SLM is added. In Section 5, we discuss the implications of this work, propose future lines of research, and discuss potential applications. Finally, we conclude our work in Section 6.

2. Conventional diffraction-limited acquisition

We consider the problem of imaging an object that emits spatially incoherent light, using a conventional diffraction-limited digital optical device. Let Σ_o denote the object plane located at distance z_o from the front principal plane of the device (see Fig. 1). We assume that the object and image planes are both perpendicular to the optical axis. The light emitted by the object is recorded by a sensor array of P square pixels, which results in a digital image $\boldsymbol{g} \in \mathbb{R}^P$. This digital image is obtained by sampling the light intensity within the image plane $g(\mathbf{x})$, $\mathbf{x} \in \Sigma_i$. Mathematically, we can model the sampling operator $\mathcal{S}$ by

$$\boldsymbol{g} = \mathcal{S}\{g\} = [(g * \phi)(\mathbf{x}_1), \ldots, (g * \phi)(\mathbf{x}_P)]^\top \Delta_t, \tag{1}$$

where $\phi(\mathbf{x})$, $\mathbf{x} \in \Sigma_i$, is an instrument response function that represents the integration effect of the sensor pixels, Δ_t is the acquisition time, $\{\mathbf{x}_p\}_{1 \leq p \leq P}$ are the centers of the pixels in the image plane Σ_i, and $*$ is the continuous-domain convolution.

The intensity g, and hence the digital image $\boldsymbol{g}$, suffers from diffraction within the imaging device. On the basis of a centered diffraction-limited model, the diffraction can be modeled through a low-pass operator $\mathcal{D}$ [20, p. 130], such that

$$g = \mathcal{D}\{f_i\} = f_i * h, \tag{2}$$

where $f_i(\mathbf{x})$, $\mathbf{x} \in \Sigma_i$, is a diffraction-free image, and $h(\mathbf{x})$, $\mathbf{x} \in \Sigma_i$ is a point-spread function corresponding to the Fraunhofer diffraction pattern of the exit pupil of the imaging system. The image f_i relates to the object-plane intensity profile f_o through geometrical optics and the inverse-square law, which depend on z_o and the type of optical system. More specifically,

$$f_i(\mathbf{x}) = \alpha f_o(M_o \mathbf{x}), \quad \mathbf{x} \in \Sigma_i \tag{3}$$

where M_o is a magnification factor, and where α is a multiplicative constant under the approximation of small angles. Inserting Equations (2) and (3) into Equation (1) yields the forward model

$$\boldsymbol{g} = \alpha \mathcal{S} \mathcal{D}\{f_o(M_o \cdot)\}, \tag{4}$$

where $\cdot$ is a dummy variable. Finally, we assume a Poissonian-Gaussian noise model [21], which yields the noisy acquisition

$$\boldsymbol{g}^\delta \sim \mathcal{P}(\boldsymbol{g} + \boldsymbol{d}) + \mathcal{N}(\boldsymbol{\mu}, \sigma_{\text{read}}^2), \tag{5}$$

where $\mathcal{P}$ and $\mathcal{N}$ are the Poisson and Gaussian distributions, respectively, $\boldsymbol{d}$ is the dark current and contributions from ambient light, $\boldsymbol{\mu}$ is an offset imposed by the sensor, and σ_{read}^2 models both the readout and the quantization noise.
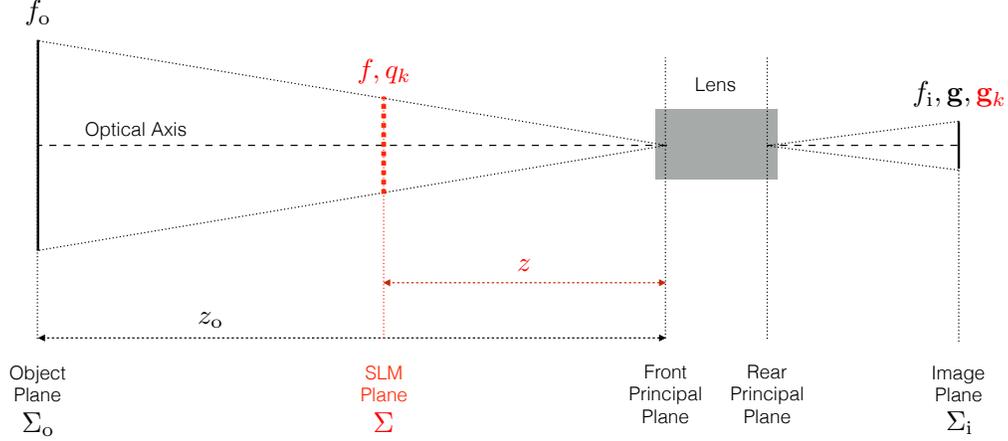
Fig. 1: Conventional approach (black) and proposed approach (red additions). The object f_o that lies on the object plane Σ_o is imaged using a conventional optical device with a lens. The numerical image g is formed within the image plane Σ_i. Compared to f_i, the ideal geometrical image of f_o, the measured image g suffers from not only sampling, but also diffraction. To circumvent both problems, we place a spatial light modulator (SLM) in the plane Σ. From a sequence of modulated images $\{g_k\}_{1 \leq k \leq K}$ acquired using some SLM patterns $\{q_k\}_{1 \leq k \leq K}$, we can recover f, which is the geometrical projection of f_o onto Σ. The dimensions are not to scale, for clarity.

Our goal is to recover a degradation-free image from the degradation-sensitive measurements, which is particularly relevant when the sampling or the diffraction effects—or both—dramatically limit the spatial resolution of a raw image g^δ. Contrary to post-processing approaches that invert Equations (4)–(5) numerically, which is prone to artifacts due the ill-posedness of the operators $\mathcal{S}$ and $\mathcal{D}$, we propose to alter the acquisition chain upfront, such that the degradation effects are neutralized.

3. Proposed diffraction-unlimited approach

3.1. Concept

We propose to make the acquisition chain described in Section 2 robust to the loss of resolution due to $\mathcal{S}$ and $\mathcal{D}$, by performing multiple acquisitions and to recover the degradation-free image using a straightforward numerical inversion after preprocessing a set of modulated images.

Our acquisition approach consists of measuring a set of dot products $\{v_k\}_{1 \leq k \leq K}$ between the image and some SLM patterns $\{q_k(\mathbf{x})\}_{1 \leq k \leq K}$, $\mathbf{x} \in \Sigma$. This is realized through the addition of a SLM to the pre-existing acquisition set-up. The key specificity of our approach is to ensure that the SLM patterns modulate the image *before* degradation occurs due to $\mathcal{S}$ and $\mathcal{D}$, which ultimately makes the dot products and the corresponding numerical reconstruction robust to these effects. This is achieved by placing the SLM sufficiently far from the optical system, at a distance z perpendicular to the optical axis (see Fig. 1). Our overall acquisition and reconstruction approach involves three main steps, as illustrated in Fig. 2 and described below.

Modulated image acquisition   The use of image modulation modifies the initial conventional forward model of Equation (4). Specifically, every SLM pattern q_k leads to measurement of the
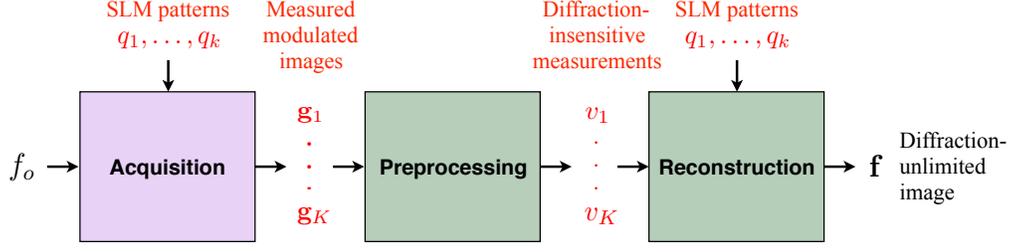
Fig. 2: Acquisition and reconstruction pipeline of the proposed approach. The set of digital modulated images $\{g_k\}_{1 \le k \le K}$ is first acquired from the object f_o using a set of SLM patterns $\{q_k\}_{1 \le k \le K}$. Then, these images are preprocessed, which leads to the diffraction-insensitive scalar measurements $\{v_k\}_{1 \le k \le K}$. Finally, the digital diffraction-free image f is reconstructed from these diffraction-insensitive measurements through a straightforward numerical inversion, based on the knowledge of the SLM patterns used for acquisition. The preprocessing step corresponds to Equation (7), the reconstruction step to Equation (10), the superscript δ being used to denote noisy measurements and estimates.

(digital) modulated image g_k given by

$$g_k = \mathcal{SD}\{q_k(M\cdot)f(M\cdot)\}, \quad 1 \le k \le K, \tag{6}$$

where $f = \alpha f_o(z_o \cdot /z)$ is the geometrical projection of f_o onto the SLM plane, and $M = M_o z/z_o$ is the corresponding magnification factor. As f_o, the intensity profile f is free of degradation due to sampling and diffraction.

As the q_k are complex-valued, whereas light intensities are positive quantities, every g_k is obtained in practice as a linear combination of several sub-acquisitions, as detailed in Sections 3.2 and 3.3 below. Finally, according to Equation (5)—which applies to every acquisition—we only have access to noisy versions g_k^δ of the modulated images g_k.

Preprocessing Once the modulated images are acquired, we numerically integrate them over their field-of-view, to produce the scalar quantities

$$v_k^\delta = \boldsymbol{u}^\top \boldsymbol{g}_k^\delta, \quad 1 \le k \le K, \tag{7}$$

where $\boldsymbol{u} = [1 \ldots 1]^\top \in \mathbb{R}^P$. While each modulated image $\boldsymbol{g}_k^\delta$ is altered by sampling and diffraction in Equation (6), the scalar quantity v_k^δ is *not* affected by sampling or diffraction. Indeed, as shown in the Appendices A.1 and A.2 , it is proportional on average to the dot product of the SLM pattern and the diffraction-free image, *i.e.*,

$$\mathbb{E}\left[v_k^\delta\right] \sim \langle q_k, f\rangle \tag{8}$$

where $\mathbb{E}$ is the expectation. This property, which is key to our acquisition approach, exploits the energy-preserving nature of diffraction.

Reconstruction As detailed in Appendix A.3, each measurement v_k^δ satisfies the discrete-scalar-product relation $\mathbb{E}\left[v_k^\delta\right] = \boldsymbol{q}_k^\top \boldsymbol{f}$, where $\boldsymbol{q}_k \in \mathbb{R}^N$ and $\boldsymbol{f} \in \mathbb{R}^N$ are discrete versions of the SLM pattern q_k and of the degradation-free image f, respectively. Defining $\boldsymbol{v}^\delta = [v_1^\delta \ldots v_K^\delta]^\top \in \mathbb{R}^K$, we have the linear model

$$\mathbb{E}\left[\boldsymbol{v}^\delta\right] = \boldsymbol{Q}\boldsymbol{f}, \tag{9}$$

where $\boldsymbol{Q} = [\boldsymbol{q}_1 \ldots \boldsymbol{q}_K]^\top \in \mathbb{R}^{K \times N}$ is the matrix containing the discrete SLM patterns. Therefore, we simply recover the diffraction-free image as

$$\boldsymbol{f}^\delta = \boldsymbol{Q}^+ \boldsymbol{v}^\delta, \tag{10}$$

where $\boldsymbol{Q}^+$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{Q}$. Here, the matrix $\boldsymbol{Q}$ is related to the choice of the SLM patterns, which is discussed in Section 3.2. Therefore, it is easy to choose $K = N$ patterns for which $\boldsymbol{Q}$ is invertible; hence $\boldsymbol{Q}^+ = \boldsymbol{Q}^{-1}$. This differs significantly from methods that invert Equation (4), where the discrete forward model depends on the physics and is typically ill-conditioned. When only a few SLM patterns are considered, i.e. $K < N$, Equation (10) provides the minimum ℓ_2 norm solution.

3.2. SLM patterns

Our choice for the SLM is dictated by two requirements. First, it is crucial to maximize light throughput, so as to limit the acquisition time needed to acquire low-intensity objects with acceptable signal-to-noise ratios (SNRs). Secondly, the SLM patterns need to capture the information of $\boldsymbol{f}$ into relatively few measurements, so as to decrease the number of measurements, and hence the time needed for an acquisition. For instance, choosing $\boldsymbol{Q}$ as the identity matrix corresponds to an extreme case where light is only transmitted through a single SLM pixel at a time. This complies poorly with these requirements.

In this study, we choose $\boldsymbol{Q}$ as the discrete Fourier basis, which transmits $\sim 64\%$ of the incident light flux and is known to sparsify natural images [22, 23]. For a SLM array of $N = N_1 \times N_2$ square pixels of size $\Delta \times \Delta$, this defines the discrete SLM patterns $\boldsymbol{q}_k = [q_k^1, \ldots, q_k^N]^\top$ as

$$q_k^n = \exp(-j2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \tag{11}$$

where j is the imaginary unit, $\mathbf{n}$ is the two-dimensional pixel coordinate associated with the n-th pixel of the SLM, and $\boldsymbol{\xi}_k$ is the two-dimensional spatial frequency of the k-th pattern, with $\boldsymbol{\xi}_k \in \{0, 1/N_1 \ldots, (N_1 - 1)/N_1\} \times \{0, 1/N_2 \ldots, (N_2 - 1)/N_2\}$.

This choice implies that the measurement vector $\boldsymbol{v}$ is the discrete Fourier transform (DFT) of $\boldsymbol{f}$. Therefore, the reconstruction step of Equation (10) simplifies to the performing of an inverse DFT, which also has the advantage of having rapid implementation with complexity $O(N \log N)$. The implementation of the spatial patterns $\boldsymbol{q}_k$ into the SLM is discussed in Section 3.3, below.

3.3. Differential acquisition strategy

As our physical set-up can only implement positive-valued SLM patterns, every complex-valued SLM pattern must be split into positive real-valued patterns to be programmed into the SLM. The acquisitions that use positive patterns are then recombined to obtain a complex-valued image $\boldsymbol{g}_k^\delta$. Here, we choose the four positive patterns $\boldsymbol{q}_{k,i}$, $1 \le i \le 4$, defined by

$$\begin{aligned} q_{k,1}^n &= \cos^+(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), & q_{k,2}^n &= \cos^-(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \\ q_{k,3}^n &= \sin^-(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), & q_{k,4}^n &= \sin^+(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \end{aligned} \tag{12}$$

where $(\cdot)^+$ and $(\cdot)^-$ are the projections onto the positive and negative orthants. These positive patterns are compatible with Equation (11) as $\boldsymbol{q}_k = (\boldsymbol{q}_{k,1} - \boldsymbol{q}_{k,2}) + j(\boldsymbol{q}_{k,3} - \boldsymbol{q}_{k,4})$. Each of the positive patterns $\boldsymbol{q}_{k,i}$ leads to the measuring of a distinct positive-modulated image $\boldsymbol{g}_{k,i}^\delta$. According to the forward and noise models of Equations (6) and (5), we have

$$\boldsymbol{g}_{k,i} = \mathcal{SD}\{q_{k,i}(M\cdot)f(M\cdot)\}, \tag{13}$$

and

$$\boldsymbol{g}_{k,i}^\delta \sim \mathcal{P}(\boldsymbol{g}_{k,i} + \boldsymbol{d}_k) + \mathcal{N}(\boldsymbol{\mu}, \sigma_{\text{read}}^2). \tag{14}$$

In the latter, we assume that $\boldsymbol{d}_k$ does not vary during the four sub-acquisitions.

Finally, we repeat the measurements L times and average them to decrease the noise. The modulated image $\boldsymbol{g}_k^\delta$ is computed as

$$\boldsymbol{g}_k^\delta = L^{-1} \sum_{\ell=1}^{L} (\boldsymbol{g}_{k,1,\ell}^\delta - \boldsymbol{g}_{k,2,\ell}^\delta) + \mathrm{j}(\boldsymbol{g}_{k,3,\ell}^\delta - \boldsymbol{g}_{k,4,\ell}^\delta), \tag{15}$$

where $\boldsymbol{g}_{k,i,\ell}^\delta$ is the ℓ-th modulated image acquired using the SLM pattern $\boldsymbol{q}_{k,i}$. As shown in Appendix A.1, the modulated image $\boldsymbol{g}_k^\delta$ is an unbiased estimate of $\boldsymbol{g}_k$, which includes the cancellation of additive ambient-light effects. This is crucial in our setting where the amount of light captured from an object at long distance can be very low.

3.4. Time budget and subsampling

As $\boldsymbol{f}$ is real-valued, only half of its DFT coefficients are nonredundant. Recalling that the SLM is an array of $N = N_1 \times N_2$ pixels, and assuming that N_1 and N_2 are even, the number of DFT coefficients for full acquisition is $K = N/2 + 2$. Implementing the differential approach of Equation (15), the time budget for a full acquisition is thus

$$t_{\text{full}} = (2N + 8)L\Delta_{\text{t}}. \tag{16}$$

The time budget of full acquisition can be substantial when the acquisition time per measurement is large (*e.g.*, for low-intensity objects that require Δ_{t} or L to be large), or when the image resolution N is large. One approach to decrease the time budget is to use a smaller number of L' of acquisitions per SLM pattern, which yields

$$t_{\text{sub}} = (2N + 8)L'\Delta_{\text{t}}, \quad L' < L. \tag{17}$$

Another approach is to acquire a subset of $K < N/2 + 2$ significant DFT coefficients. In this case, we consider the low-frequency diamond scheme, which yielded the best image-reconstruction results in [24]. The corresponding time budget is

$$t_{\text{sub}} = 4KL\Delta_{\text{t}}, \quad K < N/2 + 2. \tag{18}$$

We also propose an adaptive subsampling approach that can preserve higher-frequency coefficients, exploiting repeated measurements. This approach is described in detail in Appendix A.4. For each of the aforementioned subsampling methods, the sampling ratio γ is defined as

$$\gamma = t_{\text{sub}}/t_{\text{full}}. \tag{19}$$

3.5. Angular resolution

As this approach is robust to diffraction and sampling that occur after modulation, the maximum angular resolution R (in radians; the smaller the better) only depends on the SLM. Under the approximation of small angles, and neglecting noise, we have

$$R = \Delta/z, \tag{20}$$

where Δ is the SLM pixel size and z is the distance between the conventional optical device and the SLM. The combination of the original device and the SLM can be seen as a single new device where the maximum angular resolution is only parameterized by Δ and z. This is in contrast to conventional optical devices, where the maximum angular resolution is fixed and is limited by diffraction.

Δ (μm)	z (m)						
	0.5	1	2	5	10	20	50
10	4.1	2.1	1.0	0.41	0.21	0.10	0.041
20	8.3	4.1	2.1	0.83	0.41	0.21	0.083
50	21	10	5.2	2.1	1.0	0.52	0.21
100	41	21	10	4.1	2.1	1.0	0.41
200	83	41	21	8.3	4.1	2.1	0.83

Table 1: Maximum angular resolutions R in arcseconds for various SLM distances z and pixel sizes Δ, according to Equation (20). The angular resolution of the human eye is ca. 60 arcseconds [27]. For a pixel size of 50 μm, setting the SLM at 10 m yields an angular resolution of one arcsecond, which is comparable to that of a 4-inch telescope.

One way to improve the angular resolution is to decrease the SLM pixel size. The lower limit for Δ depends on the available technology (*e.g*, digital micro mirrors, translucent or reflective liquid crystal displays). Ultimately, Δ must be larger than the wavelength used for acquisition, to avoid diffraction effects during modulation [25]. For a fixed pixel size Δ, a target angular resolution R can be achieved by setting the SLM distance z accordingly, *i.e.*, choosing $z \geq \Delta/R$. For instance, for a pixel size of 50μm, setting the SLM at a distance of 10 m yields an angular resolution of $5 \cdot 10^{-6}$ radians, which is equivalent to approximately one arcsecond. This angular resolution is comparable to that of a regular 4-inch telescope [26]. In Table 1, we report more example values of R as a function of z and Δ.

4. Results

4.1. Experimental settings

We evaluate this approach considering a webcam (USB HD C270; Logitech) as the conventional optical device and the front screen of a commercial showcase (ClearVue Lite CV101LV1) as the SLM. The object is placed and illuminated inside the same showcase. Overall, the scene setting follows the configuration of Figure 1, where all of the distances and the effective SLM area ensure that the object is fully modulated and captured by our webcam according to the requirements of Section 3.

The webcam camera has a resolution of 1280×960 pixels, an angle of view of 60 degrees, and a focal length of 4 mm. It produces color images in compressed JPEG format that we converted to grayscale. According to the trigonometric relations between these quantities [28], the pixel size of the camera is 2.9μm, with a corresponding angular resolution of 135 arcseconds. Importantly, this pixel size is comparable to the theoretical optical diffraction limit associated with the webcam parameters, and can thus be used as a meaningful reference to assess the resolving power of our approach. For instance, the Airy-disk diameter $D_A = 2.44\lambda N$ obtained at average optical wavelength $\lambda = 550$ nm [29] and low focal ratio $N = 2$ is $D_A = 2.7\mu m$.

The SLM has a liquid crystal display of 1024×600 pixels with $\Delta = 210\mu m$ and a contrast ratio of 500:1. For our experiments, we only use an effective area of $N = 64 \times 64$ pixels located at the center of the liquid crystal display, with the rest of the pixels set to block out all of the light. For conventional acquisitions, we set the effective SLM region of 64×64 pixels to transmit all of the light, while leaving the rest of the SLM pixels set to block out all of the light. This ensures

that the same object field-of-view is acquired with and without modulation.

The object is the letter "T" as the black capital (width 9.5 mm, height 8 mm), as shown in Figure 3a. The object is placed inside the showcase, which leads to $z = 3.80$ m and $z_o = 3.87$ m, respectively. As viewed from the webcam, the effective size of f_o captured by the SLM is thus 730×730 arcseconds, or as the equivalent, 5.4×5.4 sensor pixels.

In our experiments, we acquire each pattern $L = 50$ times during $\Delta_t = 42$ ms. The object illumination, SLM transmittivity gain, and webcam gain are set to maximize the brightness, while avoiding saturation of the webcam. We perform the acquisition in a dark room, to minimize the influence of variations in ambient light during acqusion. To ensure the correct synchronization between the acquisitions performed by the webcam and the generation of the SLM patterns, a latency of 0.5 s is added between the successive measurements. As this latency is only relevant to our particular software implementation, it is not included in our time budget.

Both the optical and the SLM devices are connected to a laptop computer (MacBook Pro; 2.4 GHz Intel Core i7; 6 GB memory). All of the acquisition and reconstruction methods are implemented in Matlab.

4.2. Proposed paradigm versus conventional acquisition

In this first experiment, we assess the proposed paradigm, and compare it with conventional acquisition where the object is acquired from the same location but without SLM modulation. Based on the practical set-up and its parameters described in Section 4.1, we acquire object coefficients and reconstruct f via Equation (10). The results of this experiment are shown in Figure 3.

In the conventional acquisition setting (Fig. 3b), the object appears in a very small central region of the acquired image, where the horizontal line at the top is due to light leaking from the showcase. When inspecting a magnified central area of 8×8 pixels (Fig. 3c), no object features can be identified, and the image is blurred. This confirms that the native resolution of the webcam is insufficient to image the object. It is worth noting that in this magnified image, the area of the bright profile is consistent with the expected size of f_o, 5.4×5.4 sensor pixels, which is derived in Section 4.1 from the dimensions of this set-up, with small differences due to diffraction and instrument response.

In the proposed-paradigm setting (Fig. 3d), the reconstructed object can be resolved and appears to be consistent with the original profile (Fig. 3a). The reconstruction also contains details that are significantly smaller than the pixel resolution and diffraction limit in the classical setting (Fig. 3c). This result demonstrates that our joint acquisition and reconstruction paradigm can image objects at a resolution that significantly exceeds the native limits of the conventional device it is built from.

As mentioned in the previous sections, one major caveat of our approach is its acquisition time. In that regard, Equation (16) implies that the time budget to acquire f in our set-up is $t_{full} = 17\,220$ s (4.8 h). In the next experiment, we thus investigate how this time budget can be mitigated while maintaining acceptable reconstruction quality.

4.3. Subsampling

In this second experiment, we investigate whether the subsampling can maintain reconstruction quality under acquisition-time budgets than are lower than the one of Section 4.2. To do so, we compare the performance of the subsampling strategies proposed in Section 3.4. For convenience, our images are reconstructed retrospectively, based on the full set of DFT coefficients, as in [30]. Each subsampling method is evaluated in terms of the SNR using the fully sampled result of Section 4.2 as reference.

First, we consider the case where $\gamma = 1/4$, $i.e.$, a four-fold reduction in the acquisition time. The images reconstructed here are shown in Figure 4. Repetition subsampling (Fig. 4a) yields
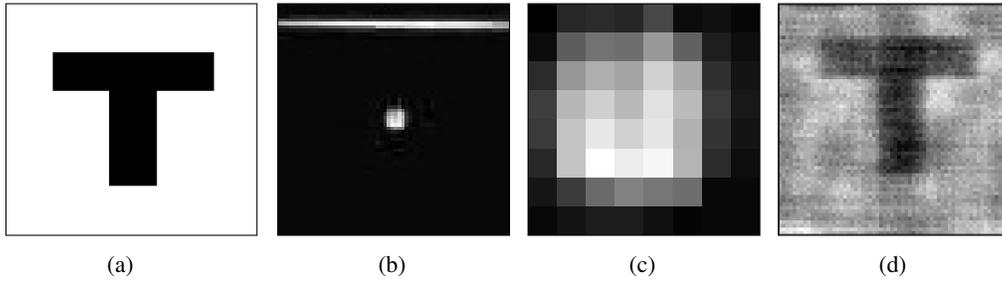
Fig. 3: The proposed paradigm compared to conventional object acquisition. (a) Reference object profile. (b) Image from the conventional optical device. (c) Magnified central area of 8×8 pixels of (b). (d) Image reconstructed using our diffraction-unlimited approach; $N = 64 \times 64$, $K = 2050$.
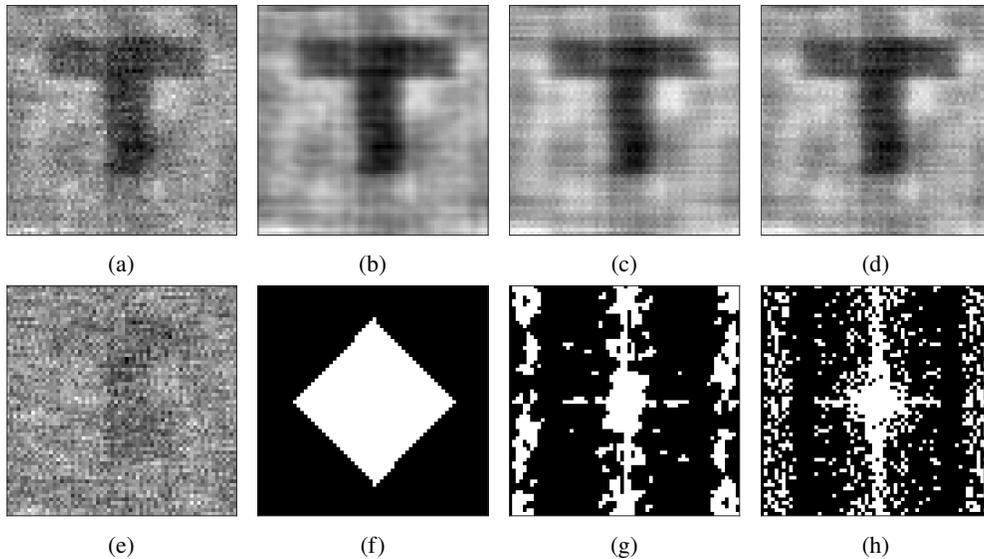

Fig. 4: Comparison between different subsampling methods. (a-d) Reconstructions for each method with a limited time budget of 1.2 h ($\gamma = 1/4$). (a) Repetition subsampling (SNR = 12.37 dB). (b) Diamond subampling (SNR = 17.87 dB). (c) Adaptive subsampling (SNR = 18.59 dB). (d) Oracle subsampling (SNR = 22.22 dB). (e) Reconstruction in the extreme case of a single repetition ($\gamma = 1/50$) for repetition subsampling (SNR = −0.12 dB). (f-h) Acquired DFT coefficients (white pixels with redundant coefficients also shown) associated with the results of (b-d). (f) Diamond subampling. (g) Adaptive subsampling. (h) Oracle subsampling.

the worst result, due to noise. The noise issue is further exacerbated in the extreme case where the image is reconstructed from a single repetition (Fig. 4e), which illustrates the need for repeat measurements. Compared to both non-adaptive approaches (Fig. 4a, b), the proposed adaptive subsampling scheme (Fig. 4c) yields the best result. This last is closest to the ideal oracle subsampling (Fig. 4d), where the highest-energy coefficients are determined and selected *a posteriori* from the full acquisition ($L = 50$; $K = N/2 + 2$). Reconstruction from adaptive sampling is also less blurry than from nonadaptive sampling, which confirms the potential of adaptive approaches to better preserve high-frequency information. The sampling pattern of
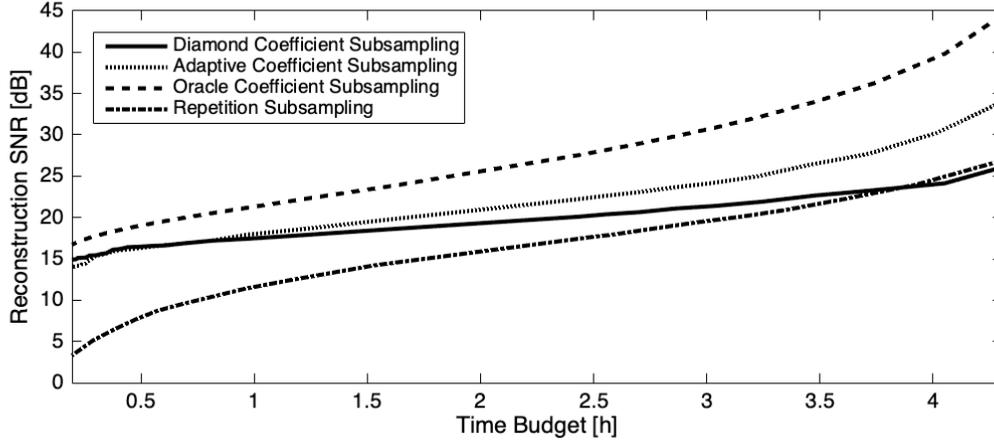
Fig. 5: Performance of the different subsampling methods for the various time budgets. The signal-to-noise ratio (in decibels) is plotted for each method as a function of the time budget in hours, varying γ accordingly.

the adaptive scheme (Fig. 4g) is also close to that of the oracle (Fig. 4h), as opposed to the nonadaptive case (Fig. 4f).

Figure 5 illustrates the SNR of the image reconstructed with different subsampling strategies with increasing sampling ratios. Adaptive sampling consistently outperforms its nonadaptive counterpart over a large range of sampling ratios. Its drop in performance at low sampling ratios is due to the constant time-budget overhead that is used to determine the coefficients relevant for repetition, as detailed in Appendix A.4. Overall, these results highlight the critical impact of the subsampling strategy and the potential of adaptive methods to preserve reconstruction quality with smaller time budgets.

5. Discussion

Our experiments confirm that the proposed paradigm can be used to resolve objects beyond the native capabilities of an optical device. It is worth emphasizing that in these experiments, the classical resolution benchmark that is outperformed by this method is the pixel size of the webcam, which is comparable to an ideal diffraction limit, as mentioned in Section 4. The actual resolving power of the webcam is even lower than this benchmark due to device nonidealities, such as optical aberrations, which also account for the blur observed in Figure 3c.

Our paradigm potentially extends to a relatively broad class of optical devices and applications. In this regard, it is important to note that the only property of the point-spread function h that is exploited to derive Equation (8) is its preservation of the total light energy, as shown in Appendix A.2. For this reason, Equation (8) also holds in the presence of optical aberrations, which can be seen as phase components of a *generalized pupil function* [20, p. 145]. Our acquisition approach would also remain valid if h were not isoplanatic, or if the light emitted from the object were spatially coherent, in which case diffraction effects would become nonlinear in intensity but remain energy preserving.

While the strength of the current study is its broad applicability to a large set of optical devices, dedicated optics can also be envisaged, drawing from previous studies on one-pixel cameras, for instance. For current single-pixel cameras, one major difference is that the SLM is primarily meant to compensate for the lack of multiple sensors, and it is integrated into the device itself instead of being placed externally. The achievable resolution is thus currently determined by the

diffraction limit of the device [24], as opposed to our acquisition approach where modulation from the SLM occurs before diffraction. However, our approach described in Section 3 is computationally similar to what is used in single-pixel cameras. In particular, both involve the acquisition of scalar products with the object using a SLM. In further work, the single-pixel architecture can thus be adapted to implement our approach, and its single sensor could physically replace the numerical-integration process in Equation (7).

In terms of applications, our paradigm can be used to either better resolve objects – as in our experiments here – or to better track them. The proposed approach can also be exploited to image objects where the distance is out of reach, for either practical or technical reasons, *e.g.*, at infinity. In that regime, the ability of our method to increase angular resolution, or, equivalently, to increase spatial resolution independent of the object distance becomes key, as there is no way to access, illuminate, or increase the size of the object in the image field of view. This is typically the case in astronomical imaging, in which case an additional SLM placed at sufficient distance inside or outside the Earth's atmosphere might further enhance the resolution limits of an existing ground or space-based telescope. Such a configuration could borrow from the *external occulter* concept [31], except that the SLM would not only allow to mask the light from unwanted sources, but also to modulate the light from the object of interest and reconstruct a more detailed image, based on our paradigm. For astronomical applications, the robustness of our paradigm to diffraction and the use of differential measurements might also prove useful, to mitigate the effects of atmospheric seeing and turbulence, depending on the relative location of the SLM. This remains the topic of further investigations.

An important limiting factor of our approach is the acquisition time required to produce a suitable reconstruction, which is caused by the very low amounts of light that can be collected by the sensor array. However, we show how particular acquisition strategies can mitigate this issue to some extent. In further work, the proposed acquisition and reconstruction methods can be adapted to decrease the time budget, to maximize the reconstruction quality, and to operate in more complex settings. For instance, satisfactory reconstructions might be obtained from fewer coefficients based on the compressed-sensing framework, assuming wavelet, total-variation, or nonlocal image priors [1, 2, 32]. Recent advances in deep learning for inverse problems might also be of benefit to our paradigm. Indeed, image reconstruction based on neural networks is intrinsically faster than compressed-sensing-based iterative algorithms, and provides improved reconstruction quality by learning image features during an offline training phase [33].

Adaptive-acquisition methods that are more advanced than the one proposed in Section 3.4 can also be developed to limit the acquisition time. Such methods might avoid repetitions of the same SLM pattern by producing specifically optimized patterns for each new measurement, based on information on signal and noise properties that would be gathered from all of the previous measurements. Furthermore, acquisition settings where the object is moving or changing might be handled by adapting our acquisition model and our algorithms.

6. Conclusion

We here propose and demonstrate experimentally a novel imaging paradigm where an optical device can be used in conjunction with a SLM to acquire and resolve remote objects with resolution that exceeds the diffraction limit of the optical device. Our acquisition strategy can be seen as the transformation of the optical device that produces degraded measurements—due to diffraction and instrument response—into a new device that produces compressed measurements. Then, the image of the scene is reconstructed by solving a simple inverse problem. Our experiments represent the first proof-of-concept that the loss of spatial and angular resolution that is intrinsic to diffraction can be circumvented through the use of specific acquisition and reconstruction strategies.

A. Appendix

A.1. Unbiasedness of the measurements

By linearity of the expectation, we obtain, from the differential expression of $\boldsymbol{g}_k^\delta$ given by Equation (15),

$$\mathbb{E}\left[\boldsymbol{g}_k^\delta\right] = (\mathbb{E}\left[\boldsymbol{g}_{k,1}^\delta\right] - \mathbb{E}\left[\boldsymbol{g}_{k,2}^\delta\right]) + \mathrm{j}(\mathbb{E}\left[\boldsymbol{g}_{k,3}^\delta\right] - \mathbb{E}\left[\boldsymbol{g}_{k,4}^\delta\right]). \tag{21}$$

Under the noise model of Equation (14), the additive biases, which include the effects of ambient light, cancel out mathematically as

$$\mathbb{E}\left[\boldsymbol{g}_k^\delta\right] = (\boldsymbol{g}_{k,1} - \boldsymbol{g}_{k,2}) + \mathrm{j}(\boldsymbol{g}_{k,3}^\delta - \boldsymbol{g}_{k,4}^\delta). \tag{22}$$

Using the forward model of Equation (13), we then obtain

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{g}_k^\delta\right] = &(\mathcal{S}\mathcal{D}\{q_{k,1}(M\cdot)f(M\cdot)\} - \mathcal{S}\mathcal{D}\{q_{k,2}(M\cdot)f(M\cdot)\}) \\
&+ \mathrm{j}(\mathcal{S}\mathcal{D}\{q_{k,3}(M\cdot)f(M\cdot)\} - \mathcal{S}\mathcal{D}\{q_{k,4}(M\cdot)f(M\cdot)\}),
\end{aligned} \tag{23}$$

which simplifies, by linearity of both $\mathcal{S}$ and $\mathcal{D}$, to

$$\mathbb{E}\left[\boldsymbol{g}_k^\delta\right] = \mathcal{S}\mathcal{D}\{q_k(M\cdot)f(M\cdot)\} = \boldsymbol{g}_k. \tag{24}$$

A.2. Robustness to diffraction and sampling

We now demonstrate that v_k^δ are robust to diffraction and sampling. In more detail, they provide inner products between the degradation-free image f and the modulating patterns q_k, up to a multiplicative constant. By linearity of the expectation, we have $\mathbb{E}\left[v_k^\delta\right] = \mathbb{E}\left[\boldsymbol{u}^\top \boldsymbol{g}_k^\delta\right] = \boldsymbol{u}^\top \mathbb{E}\left[\boldsymbol{g}_k^\delta\right]$. Therefore, as measurements are unbiased as shown in Equation (24), we have $\mathbb{E}\left[v_k^\delta\right] = \boldsymbol{u}^\top \boldsymbol{g}_k$. By definition of $\boldsymbol{g}_k$, we have

$$\mathbb{E}\left[v_k^\delta\right] = \boldsymbol{u}^\top \mathcal{S}\mathcal{D}\{q_k(M\cdot)f(M\cdot)\}. \tag{25}$$

Substituting the sampling operator $\mathcal{S}$ with its definitions leads to

$$\begin{aligned}
\mathbb{E}\left[v_k^\delta\right] &= \Delta_\mathrm{t} \sum_p \left(\mathcal{D}\{q_k(M\cdot)f(M\cdot)\} * \phi\right)(\mathbf{x}_p) \\
&= \Delta_\mathrm{t} \sum_p \int_{\Sigma_\mathrm{i}} \mathcal{D}\{q_k(M\cdot)f(M\cdot)\}(\mathbf{x})\,\phi(\mathbf{x}_p - \mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned} \tag{26}$$

Assuming that the pixel instrument function $\phi(\mathbf{x}_p - \mathbf{x})$ indicates the location of the p-th pixel of the sensor, the instrument function forms a partition of unity, *i.e.*, $\sum_p \phi(\mathbf{x}_p - \mathbf{x}) = 1$, and the convolution simplifies to

$$\mathbb{E}\left[v_k^\delta\right] = \Delta_\mathrm{t} \int_{\Sigma_\mathrm{i}} \mathcal{D}\{q_k(M\cdot)f(M\cdot)\}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{27}$$

In a similar fashion, the convolution kernel h satisfies $\int h(\mathbf{x})\mathrm{d}\mathbf{x} = 1$ because diffraction preserves the total light energy. Therefore, assuming that the diffracted modulated images $\mathcal{D}\{q_k(M\cdot)f(M\cdot)\}$ lie within the field of view of the device, we obtain

$$\mathbb{E}\left[v_k^\delta\right] = \Delta_\mathrm{t} M^{-2} \int_\Sigma q_k(\mathbf{x})f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \Delta_\mathrm{t} M^{-2} \langle q_k, f\rangle, \tag{28}$$

which is independent of $\mathcal{S}$ and $\mathcal{D}$.

A.3. Discretization

Each modulation pattern $q_k(\mathbf{x})$ is implemented using a vector $\boldsymbol{q}_k = [q_k^1, \ldots, q_k^N]^\top$ that indicates the value of each of the SLM pixels. Mathematically, we have

$$q_k(\mathbf{x}) = \sum_n q_k^n b(\mathbf{x} - \mathbf{x}_n), \tag{29}$$

where $b(\mathbf{x})$ is a square box function of size Δ that represents the shape of the SLM pixels. Therefore Equation (28) expands as

$$\mathbb{E}\left[v_k^\delta\right] = \Delta_t M^{-2} \int_\Sigma \sum_n q_k^n b(\mathbf{x} - \mathbf{x}_n) f(\mathbf{x}) \, d\mathbf{x} = \Delta_t M^{-2} \sum_n q_k^n \left(\int_\Sigma f(\mathbf{x}) b(\mathbf{x} - \mathbf{x}_n) \, d\mathbf{x} \right). \tag{30}$$

Defining $\boldsymbol{f} = [f^1, \ldots, f^N]^\top$ as the discrete version of $f(\mathbf{x})$ that matches the resolution and pixel size of the SLM, *i.e.*,

$$f^n = \Delta_t M^{-2} \int_\Sigma f(\mathbf{x}) b(\mathbf{x} - \mathbf{x}_n) d\mathbf{x}, \tag{31}$$

we obtain the discrete-scalar-product relation

$$\mathbb{E}\left[v_k^\delta\right] = \sum_{n=1}^N q_k^n f^n = \boldsymbol{q}_k^\top \boldsymbol{f}. \tag{32}$$

A.4. Adaptive subsampling

Subsampling methods such as the low-frequency diamond scheme proposed in [24] select coefficients *a priori*, regardless of image properties. Therefore, such methods can miss relevant high frequencies. Inspired by [30], we propose an adaptive subsampling scheme that mitigates this problem. Specifically, we propose to first make a quick but exhaustive acquisition of all of the coefficients, and then to repeat the acquisition of the highest coefficients only, to increase their SNR.

Retaining K significant coefficients and acquiring the significant coefficients L times at most, we obtain the adaptive measurement vector as follows:

1. Acquire all patterns, *i.e.*, $\{\boldsymbol{q}_k\}$ for $1 \leq k \leq N$. The resulting measurement vector is denoted by $\boldsymbol{v}^{\delta,1}$

2. Low-pass filter the image of the first-pass measurements $\boldsymbol{v}^{\delta,1}$ with a Gaussian filter of unit variance; this yields the filtered coefficients $\hat{\boldsymbol{v}}^{\delta,1}$

3. Find the locations of the K highest (absolute) values of the filtered measurement vector $\hat{\boldsymbol{v}}^{\delta,1}$. The set of indices indicating relevant coefficients is denoted by Ω.

4. Acquire the relevant patterns, *i.e.*, $\{\boldsymbol{q}_k\}_{k\in\Omega}$, $L-1$ more times. The resulting measurement vectors are denoted by $\boldsymbol{v}^{\delta,\ell}$, $2 \leq \ell \leq L$, where the coefficients that are not acquired are set to zero. The first-pass acquisition is cleaned by setting $v_k^{\delta,1}$ to zero for all $k \notin \Omega$.

5. Average the measurement vectors to get the adaptive measurement vector $\boldsymbol{v}^\delta$. Mathematically, $\boldsymbol{v}^\delta = \frac{1}{L} \sum_{\ell=1}^L \boldsymbol{v}^{\delta,\ell}$.

The low-pass filtering in Step 2 and the elimination of nonrelevant coefficients in Step 4 both aim to reduce the effect of noise on the initial single pass. The time budget associated with this adaptive scheme is

$$t_{\text{sub}} = (4KL + 2N + 8 - 4K)\Delta_t, \quad K < N/2 + 2, \tag{33}$$

where the overhead $(2N + 8 - 4K)\Delta_t$ is due to the adaptive coefficient selection in Steps 1-3.

Disclosures

The authors declare no conflicts of interest.

References

1. E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," arXiv preprint math/0409186 (2004).
2. E. J. Candès, "Compressive sampling," in *Proceedings of the international congress of mathematicians,* vol. 3 (Madrid, Spain, 2006), pp. 1433–1452.
3. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE signal processing magazine **25**, 83–91 (2008).
4. N. Radwell, K. J. Mitchell, G. M. Gibson, M. P. Edgar, R. Bowman, and M. J. Padgett, "Single-pixel infrared and visible microscope," Optica **1**, 285–289 (2014).
5. C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla, "Terahertz compressive imaging with metamaterial spatial light modulators," Nat. Photonics **8**, 605–609 (2014).
6. Q. Pian, R. Yao, N. Sinsuebphon, and X. Intes, "Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging," Nat. Photonics **11**, 411–414 (2017).
7. F. Rousset, N. Ducros, F. Peyrin, G. Valentini, C. D'Andrea, and A. Farina, "Time-resolved multispectral imaging based on an adaptive single-pixel camera," Opt. Express **26**, 10550–10558 (2018).
8. D. J. Starling and J. Ranalli, "Compressive sensing for spatial and spectral flame diagnostics," Sci. Reports **8**, 2556 (2018).
9. R. Horisaki, H. Matsui, and J. Tanida, "Single-pixel compressive diffractive imaging with structured illumination." Appl. optics **56**, 4085–4089 (2017).
10. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," Nat. Photonics **13**, 13–20 (2019).
11. L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," The J. Cell Biol. **190**, 165–175 (2010).
12. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Opt. Lett. **19**, 780–782 (1994).
13. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science **313**, 1642–1645 (2006).
14. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," Nat. Methods **3**, 793 (2006).
15. V. Micó, Z. Zalevsky, C. Ferreira, and J. García, "Superresolution digital holographic microscopy for three-dimensional samples," Opt. Express **16**, 19260–19270 (2008).
16. T. Gutzler, T. R. Hillman, S. A. Alexandrov, and D. D. Sampson, "Coherent aperture-synthesis, wide-field, high-resolution holographic microscopy of biological tissue," Opt. Lett. **35**, 1136–1138 (2010).
17. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution fourier ptychographic microscopy," Nat. Photonics **7**, 739 (2013).
18. J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," Publ. Astron. Soc. Pac. **114**, 1051–1069 (2002).
19. Puschmann, K. G. and Kneer, F., "On super-resolution in astronomical imaging," Astron. & Astrophys. **436**, 373–378 (2005).
20. J. W. Goodman, *Introduction to Fourier optics (2nd edition)* (McGraw Hill Higher Education, 1996).
21. A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," IEEE Transactions on Image Process. **17**, 1737–1754 (2008).
22. D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," Ophthalmic Physiol. Opt. **12**, 229–232 (1992).
23. A. Torralba and A. Oliva, "Statistics of natural image categories," Network: computation neural systems **14**, 391–412 (2003).

24. Z. Zhang, X. Wang, G. Zheng, and J. Zhong, "Hadamard single-pixel imaging versus fourier single-pixel imaging," Opt. Express **25**, 19619–19639 (2017).

25. K. M. Schmitt and M. Rahm, "Evaluation of the impact of diffraction on image reconstruction in single-pixel imaging systems," Opt. Express **24**, 23863–23871 (2016).

26. M. Mobberley, *Astronomical equipment for amateurs* (Springer Science & Business Media, 1999).

27. M. Yanoff and J. S. Duker, *Ophthalmology, Maryland Heights, MO* (Elsevier, 2009).

28. E. McCollough, "Photographic topography," Ind. A Mon. Mag. Devoted to Sci. Eng. Mech. Arts p. 65 (1893).

29. E. R. Fossum, "What to do with sub-diffraction-limit (SDL) pixels?-A proposal for a gigapixel digital film sensor (DFS)," in *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors,* (2005), pp. 214–217.

30. F. Rousset, N. Ducros, A. Farina, G. Valentini, C. D'Andrea, and F. Peyrin, "Adaptive basis scan by wavelet prediction for single-pixel imaging," IEEE Transactions on Comput. Imaging **3**, 36–46 (2017).

31. W. A. Traub and B. R. Oppenheimer, "Direct imaging of exoplanets," Exoplanets pp. 111–156 (2010).

32. G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *European Conference on Computer Vision,* (Springer, 2008), pp. 57–68.

33. C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, "Deep learning for real-time single-pixel video," Sci. reports **8**, 2369 (2018).