Diffraction-Unlimited Imaging Based on Conventional Optical Devices

Nicolas Ducros, Aurélien Bourquard

Diffraction-Unlimited Imaging Based on Conventional Optical Devices

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

[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*
[*] *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 (SLM) at 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 the ability of our approach to 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.

References

1. E. 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 *et al.*, "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. Yanoff and J. S. Duker, *Ophthalmology, Maryland Heights, MO* (Elsevier, 2009).
27. E. McCollough, "Photographic topography," Ind. A Mon. Mag. Devoted to Sci. Eng. Mech. Arts p. 65 (1893).
28. 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.
29. 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).
30. G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *European Conference on Computer Vision,* (Springer, 2008), pp. 57–68.
31. 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).
32. W. A. Traub and B. R. Oppenheimer, "Direct imaging of exoplanets," Exoplanets pp. 111–156 (2010).

1. Introduction

The resolution of an imaging system, *i.e.*, its ability to separate points that are located at small angular position, is limited by the density of sensors and by diffraction. Thanks to advances in signal processing and computer science, computational approaches have been proposed to address cases where the amount of sensors is the limiting factor. In particular, the compressed-sensing framework allows to reconstruct images from substantially less sensors than according to the Shannon-Nyquist sampling limit [1, 2]. A practical implementation of compressed sensing is 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 on this topic. In parallel to this trend, however, the spatial density of sensors has increased substantially, with native camera resolutions in mere consumer mobile phones now exceeding ten megapixels. In most consumer and professional optical devices, computational methods that compensate for 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 diffraction occurring inside the optical system. This physical limit has been extensively addressed in fluorescence microscopy. With the advent of super-resolution techniques, different technologies, each with several variants, are now available [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] or stochastic optical reconstruction microscopy [14]), and 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], and Fourier ptychography [17]). However, when the object is too far to be accessed (*e.g.*, in remote sensing or astronomy), none of the latter approaches apply. In this case, only post-processing approaches can be implemented to deconvolve the instrument response [18]. While deconvolution may 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 structures [19].

In this paper, we propose a computational imaging approach that allows objects that are too far to be illuminated or accessed to be imaged beyond the diffraction limit. 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, adding a spatial light modulator (SLM) at some distance from the optical device. After the acquisition of a sequence of images for different SLM patterns, the object can be reconstructed by a simple procedure. To the best of our knowledge, this is the first time that a practical setup based on these techniques is 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 our 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 a practical optical implementation that is based on a webcam, comparing its results with respect to the conventional case where no SLM is added. In Section 5, we discuss the implications of our work, propose future lines of research, and discuss potential applications.

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, resulting in a digital image $\boldsymbol{g} \in \mathbb{R}^P$. The latter is obtained by sampling the light intensity within the image plane $g(\mathbf{x})$, $\mathbf{x} \in \Sigma_\mathrm{i}$. Mathematically, we can model the sampling operator $\mathcal{S}$ by

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

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

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

$$g = \mathcal{D}\{f_\mathrm{i}\} = f_\mathrm{i} * h \tag{2}$$

where $f_\mathrm{i}(\mathbf{x})$, $\mathbf{x} \in \Sigma_\mathrm{i}$, is a diffraction-free image and $h(\mathbf{x})$, $\mathbf{x} \in \Sigma_\mathrm{i}$ is 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, depending on z_o and on the type of optical system. More specifically,

$$f_\mathrm{i}(\mathbf{x}) = \alpha f_\mathrm{o}(M_\mathrm{o}\mathbf{x}), \quad \mathbf{x} \in \Sigma_\mathrm{i} \tag{3}$$

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

$$\boldsymbol{g} = \alpha \mathcal{S}\mathcal{D}\{f_\mathrm{o}(M_\mathrm{o}\cdot)\}, \tag{4}$$

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

$$\boldsymbol{g}^\delta = \mathcal{P}(\boldsymbol{g} + \boldsymbol{d}) + \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2_\mathrm{read}), \tag{5}$$

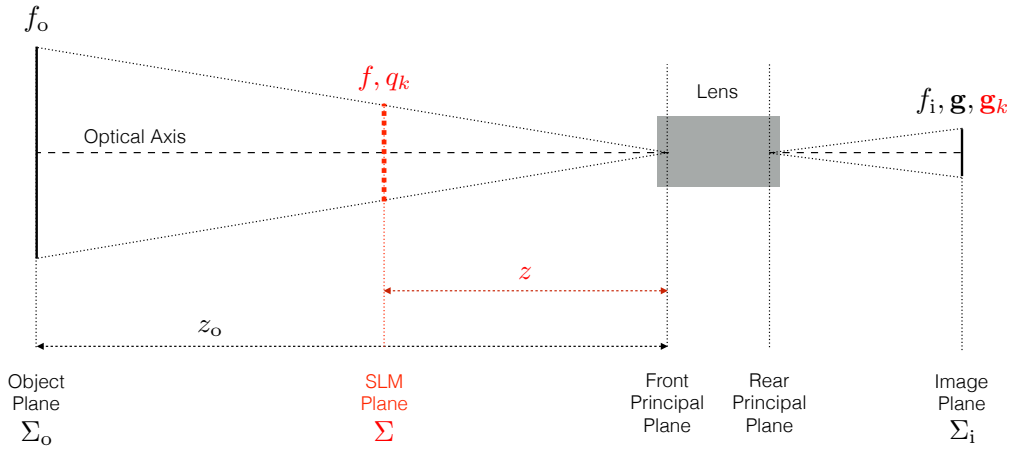


Fig. 1: Conventional approach (black) and proposed approach (differences in red). The object f_o lying in the object plane Σ_o is imaged using a conventional optical device that comprises a lens. The numerical image $\boldsymbol{g}$ is formed within the image plane Σ_i. Compared to f_i, the ideal geometrical image of f_o, the measured image $\boldsymbol{g}$ suffers from sampling and diffraction. To circumvent this problem, we place a spatial light modulator (SLM) in the plane Σ. Using some SLM patterns q_k, $1 \leq k \leq K$ for acquisition and post-processing, the modulated images $\boldsymbol{g}_k$ allow to recover f, the geometrical projection of f_o onto Σ. The dimensions are not to scale for clarity.

where $\mathcal{P}$ and $\mathcal{N}$ denotes the Poisson and Gaussian distributions, $\boldsymbol{d}$ represents the dark current and contributions from ambient light, $\boldsymbol{\mu}$ is an offset imposed by the sensor, and $\boldsymbol{\sigma}^2_{\text{read}}$ models both the readout and quantization noise.

Our goal is to recover a degradation-free image from the degradation-sensitive measurements, which is particularly relevant when sampling or diffraction effects—or both—dramatically limit the spatial resolution of raw image $\boldsymbol{g}^\delta$. Contrary to post-processing approaches that invert (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 by a straightforward numerical inversion after preprocessing a set of modulated images.

Our acquisition approach consists in measuring a set of K dot products v_k between the image and corresponding complex-valued spatial patterns $q_k(\mathbf{x})$, $\mathbf{x} \in \Sigma$, $1 \leq k \leq K$. This is realized through the addition of a SLM to the pre-existing acquisition setup. The key specificity of our approach is to ensure that the q_k 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 outside the optical system, at a distance z perpendicular to the optical axis (see Fig. 1). Our overall acquisition and reconstruction approach involves the three main steps described below.

Modulated image acquisition The use of image modulation modifies the initial conventional forward model of Eq. (4). Specifically, every modulating pattern q_k can be mathematically

associated with a complex-valued sampled digital image $\boldsymbol{g}_k \in \mathbb{C}^P$ of the form

$$\boldsymbol{g}_k = \mathcal{SD}\{q_k(M\cdot)f(M\cdot)\}, \tag{6}$$

where $f = \alpha f_o(z_o \cdot / z)$ is the geometrical projection of f_o onto the SLM plane, and where $M = M_o z / z_o$ is the corresponding magnification factor. Since the q_k are complex-valued whereas light intensities are positive quantities, every $\boldsymbol{g}_k$ 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 Eq. (5)—which applies to every sub-acquisition—we only have access to noisy versions $\boldsymbol{g}_k^\delta$ of the $\boldsymbol{g}_k$.

Dot product estimation Once the modulated images $\boldsymbol{g}_k^\delta$ are acquired, we numerically integrate them over their whole field of view, which produces the scalar quantities

$$v_k^\delta = \boldsymbol{u}^\top \boldsymbol{g}_k^\delta, \tag{7}$$

where $\boldsymbol{u} = [1 \dots 1]^\top \in \mathbb{R}^P$. While the content of the modulated images $\boldsymbol{g}_k^\delta$ is altered by sampling and diffraction in (6) after modulation, the spatially integrated quantities v_k^δ are not affected by sampling or diffraction, and correspond to unbiased estimates of the dot products between the q_k and the degradation-free image f, as shown in Appendices 6.1 and 6.2. This property, which is key to our acquisition approach, exploits the fact that diffraction is energy-preserving.

Discrete reconstruction Let $\boldsymbol{q}_k = [q_k^1, \dots, q_k^N]^\top$ be the sequence of N pixels defining the k-th modulating pattern $q_k(\mathbf{x})$. As explained and detailed in Appendix 6.3, our measurements v_k^δ satisfy the discrete-scalar-product relation

$$\mathbb{E}\left[v_k^\delta\right] = \boldsymbol{q}_k^\top \boldsymbol{f}, \tag{8}$$

where $\boldsymbol{f} \in \mathbb{R}^N$ is a discrete version of $f(\mathbf{x})$ whose resolution matches the one of the $\boldsymbol{q}_k$. Accordingly, defining $\boldsymbol{v}^\delta = [v_1^\delta \dots v_K^\delta]^\top$ as the measurement vector and $\boldsymbol{Q} = [\boldsymbol{q}_1 \dots \boldsymbol{q}_K]^\top$ as the matrix containing the corresponding SLM patterns, we estimate the degradation-free image $\boldsymbol{f}$ as

$$\boldsymbol{f}^\delta = \boldsymbol{Q}^+ \boldsymbol{v}^\delta, \tag{9}$$

where $\boldsymbol{Q}^+$ the Moore-Penrose pseudo-inverse of $\boldsymbol{Q}$, which corresponds to the regular matrix inverse when $\boldsymbol{Q}$ is invertible but is associated with the minimum-energy solution when the system $\boldsymbol{v} = \boldsymbol{Q}\boldsymbol{f}$ is under-determined. The matrix $\boldsymbol{Q}$ is related to the choice of the SLM patterns, which is discussed in section 3.2.

3.2. Spatial modulating patterns

Our choice for the spatial patterns q_k that are used for modulation with the SLM is dictated by two requirements. First, it is crucial to maximize light throughput efficiency, which allows to limit the acquisition time per measurement v_k^δ needed to acquire low intensity objects with acceptable signal-to-noise ratio. Second, the patterns q_k should ideally capture the energy of $\boldsymbol{f}$ into relatively few measurements, which allows to decrease the amount of measurements K 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 poorly complies with these requirements. In this work, 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 modulating patterns $\boldsymbol{q}_k$ as

$$q_k^n = \exp(-\mathrm{j}2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \tag{10}$$

where j is the imaginary unit, $\mathbf{n}$ is the 2D pixel coordinate associated with the n-th index, and $\boldsymbol{\xi}_k$ is the 2D 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 of $\boldsymbol{f}$. The reconstruction step (9) thus amounts to performing an inverse Fourier transform, which also has the advantage of having a fast 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

Since our acquisition setup only involves positive-valued light intensities, every complex-valued modulating pattern $\boldsymbol{q}_k$ must be split into four positive real-valued patterns to be programmed into the SLM, and the corresponding sub-acquisitions that are obtained then recombined to obtain the complex-valued image $\boldsymbol{g}_k^\delta$. The definition of $\boldsymbol{q}_k$ in Eq. (10) determines the discrete positive patterns $\boldsymbol{q}_{k,i}$, $1 \le i \le 4$, as

$$\begin{aligned} q_{k,1}^n &= \cos^+(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), & q_{k,2}^n &= \cos^-(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \\ q_{k,3}^n &= \sin^-(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), & q_{k,4}^n &= \sin^+(2\pi \boldsymbol{\xi}_k^\top \mathbf{n}), \end{aligned} \tag{11}$$

where $(\cdot)^+$ and $(\cdot)^-$ denote the projections onto the positive and negative orthants, and where

$$\boldsymbol{q}_k = (\boldsymbol{q}_{k,1} - \boldsymbol{q}_{k,2}) + \mathrm{j}(\boldsymbol{q}_{k,3} - \boldsymbol{q}_{k,4}). \tag{12}$$

Following our forward model, each physical SLM pattern $q_{k,i}$ that is addressed by the sequence $\boldsymbol{q}_{k,i}$ produces a distinct sub-acquisition $\boldsymbol{g}_{k,i}^\delta$. Specifically, following the forward model (6) and the noise model (5), we have

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

and

$$\boldsymbol{g}_{k,i}^\delta = \mathcal{P}(\boldsymbol{g}_{k,i} + \boldsymbol{d}_k) + \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\mathrm{read}}^2), \tag{14}$$

where we assume that $\boldsymbol{d}_k$ cannot vary between sub-acquisitions. In full generality, the same sub-acquisition $\boldsymbol{g}_{k,i}^\delta$ can be repeated L times to decrease noise, which yields acquisitions of the form $\boldsymbol{g}_{k,i,\ell}^\delta$, where ℓ is the ℓ-th acquisition with $1 \le \ell \le L$. Given Eq. (12), and by virtue of the linearity of (6), the resulting modulated image $\boldsymbol{g}_k^\delta$ is then obtained in our setup as

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

Depending on the amount of measurements K and on L, measuring $\boldsymbol{v}^\delta$ thus requires $4KL$ image acquisitions in total. As shown in Appendix 6.1, the modulated image $\boldsymbol{g}_k^\delta$ is an unbiased estimate of $\boldsymbol{g}_k$, which includes the cancellation of additive ambient-light effects. The latter is crucial in our setting where the amount of light captured from the object of interest can be low due to z_o.

3.4. Time budget and subsampling

Since $\boldsymbol{f}$ is real-valued, only half of its DFT coefficients are non-redundant. 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 a full acquisition is $K = N/2 + 2$. Implementing the differential approach of (15), the time budget for a full acquisition is thus

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

The time budget of a full acquisition can be substantial when the acquisition time per measurement v_k^δ is large (which occurs when Δ_t or L is large) or when the image resolution N is large. One approach to decrease the time budget is to use a smaller amount L' of acquisitions per SLM pattern, which yields

$$t_{\text{sub}} = (2N + 8)L'\Delta_t, \quad L' < L. \tag{17}$$

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

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

We also designed an adaptive subsampling approach that can preserve higher-frequency coefficients, exploiting our particular acquisition scheme where acquisitions are repeated with the same patterns. This approach is described in details in Appendix 6.4. 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. Maximum angular resolution

Since our acquisition approach is robust to the diffraction and sampling that occur after modulation, the maximum angular resolution R that it can achieve directly depends on the configuration of the SLM. Specifically, under the approximation of small angles, and neglecting noise,

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

with R in radians. The smaller R, the better. One way to improve the angular resolution is to decrease the SLM pixel size Δ. The lower limit for Δ depends on what is achievable with 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 fixed Δ, a target angular resolution R can be achieved by setting the SLM distance z accordingly, with $z \geq \Delta/R$. From that perspective, our approach is practically most relevant when objects are located very far from the device, with $z_o \gg z$.

Overall, the combination of the original device and the SLM shown in Fig. 1 can be seen as a single new device whose maximum angular resolution R is parameterized by Δ and z. This contrasts with conventional optical devices where R is fixed and limited by diffraction. In Table 1, we report example values of R as a function of z and Δ for our acquisition approach.

4. Experiments

4.1. Settings

Based on a practical setup, we conducted experiments to evaluate our approach and compare it with conventional acquisition. A USB Logitech HD C270 webcam was used as optical device, and the front screen of a commercial ClearVue Lite CV101LV1 showcase was used as SLM. The object was placed and illuminated inside the same showcase. Overall, the scene setting followed the configuration of Fig. 1, with all distances and the effective SLM area ensuring that 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 4mm. It produces color images in compressed JPEG format that we converted to grayscale. According to the trigonometric relations between these quantities [27], the pixel size is 2.9μm, with a corresponding angular resolution of 135 arcseconds. Importantly, this

Δ (in μm)	z (in 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 (20). The angular resolution of the human eye is ca. 60 arcseconds [26].

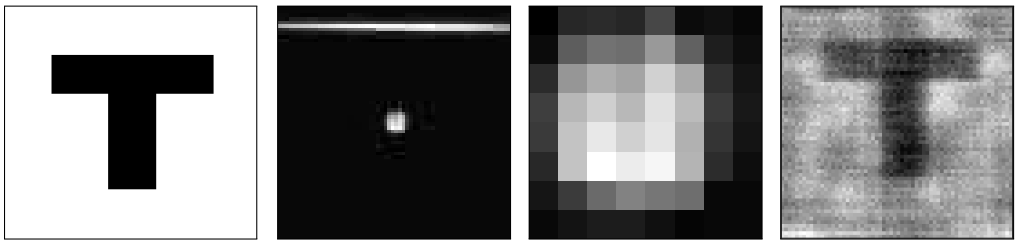
pixel size is comparable to the theoretical optical diffraction limit associated with the webcam parameters, and can thus be used as a meaningful reference to assess the resolving power of our approach. For instance, the Airy-disk diameter $D_{\mathrm{A}} = 2.44\lambda N$ obtained at average optical wavelength $\lambda = 550$nm [28] and low focal ratio $N = 2$ is $D_{\mathrm{A}} = 2.7\mu$m. This limit only constitutes an upper bound in terms of the resolving power that can be achieved in practice by the classical optical device. In particular, the webcam is composed of relatively cheap hardware components that are prone to aberrations, which introduces additional non-idealities. This implies that, if our approach can image details that are substantially smaller than the pixel size, its optical resolution definitely exceeds the one of the classical system without the SLM.

The SLM, which corresponds to the front screen of the ClearVue showcase, is a liquid crystal display (LCD) of 1024×600 pixels that modulates light transmission, with $\Delta = 210\mu$m and a contrast ratio of 500:1. For our experiments, we only used an effective area of $N = 64 \times 64$ pixels located at the center of the LCD for modulation, with the rest of the pixels set to block out all the light. In the conventional-acquisition case, we set the effective SLM region of 64×64 pixels to transmit all the light while leaving the rest of the SLM pixels set to block out all the light. This ensured that the exact same object field of view was acquired without any modulation.

The object is the black capital "T" letter shown in Fig. 2a. The letter width and height is $9.5 \times 8\mathrm{mm}^2$. Since the object was placed inside the ClearVue showcase whose front screen was also used as SLM, it had to be located relatively close to the SLM for practical reasons. Accordingly, the distances between the SLM and object planes and the device were $z = 3.80$m and $z_{\mathrm{o}} = 3.87$m, respectively. As viewed from the webcam, the effective size of f_{o} captured by the SLM was thus 730×730 arcseconds, or, equivalently, 5.4×5.4 sensor pixels.

In our experiments, we performed a total of $L = 50$ acquisitions per SLM pattern, each being associated with an acquisition time of $\Delta_{\mathrm{t}} = 42$ms. The object illumination, SLM transmittivity gain, and webcam gain were set to maximize the brightness of all acquisitions while ensuring a linear response without saturation. The effects of ambient light and of other light sources were also minimized by performing the acquisition in a closed room where all lights were shut down, except the one used for the object illumination. To ensure proper synchronization between the acquisitions performed by the webcam and the generation of the SLM patterns, a latency of 0.5s was also added between successive measurements. Since this latency was only relevant to our particular software implementation, it was not included in our time budget.

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



(a) (b) (c) (d)

Fig. 2: Proposed paradigm compared to conventional object acquisition. (a) Reference object profile. (b) Image of the object at distance z_o from the device in the conventional setting. (c) Magnified central area of 8×8 pixels from the same image. (d) Image of the object at same distance z_o reconstructed from modulated acquisitions using the proposed paradigm ($K = 2050$).

4.2. Proposed Paradigm Versus Conventional Acquisition

In this first experiment, we assess the proposed paradigm and compare it with the conventional approach where the object is acquired from the same location but without SLM modulation. Based on the practical setup and its parameters described in Section 4.1, we acquire object coefficients and reconstruct $\boldsymbol{f}$ via Eq. (9). The results of this experiment are shown in Fig. 2.

In the conventional acquisition setting (Fig. 2b), the object appears in a very small central region of the acquired image, the horizontal line at the top being due to light leaking from the showcase. When inspecting a magnified central area of 8×8 pixels (Fig. 2c), no object feature can be identified. This confirms that the native resolution of the webcam is insufficient to image the object. It is worthwhile to note 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) derived in Section 4.1 from the dimensions of our setup, up to small differences due to diffraction and instrument response.

In the case of the proposed paradigm (Fig. 2d), the reconstructed object can be resolved and appears to be consistent with the original profile (Fig. 2a). This result confirms the ability of our joint acquisition and reconstruction paradigm to image objects with higher optical resolution than according to the native capabilities of the webcam. In particular, the imaging capabilities of our paradigm lie beyond the diffraction limit of the optical system, knowing that the field of view of the reconstructed image corresponds to a sensor area of 15.7×15.7 microns (5.4×5.4 pixels) only. As mentioned in the previous sections, one major caveat of our approach is its acquisition time. In that regard, Eq. (16) implies that the time budget to acquire f in our setup is $t_{\text{full}} = 17220$s (4.8 hours). In the next experiment, we thus investigate how this time budget can be mitigated while maintaining an acceptable reconstruction quality.

4.3. Subsampling Strategies

In this second experiment, we investigated the ability of subsampling to maintain reconstruction quality under lower acquisition-time budgets than in Section 4.2. To do so, we exploited and compared the performance of the subsampling strategies proposed in Section 3.4, considering the exact same reconstruction problem. For convenience, our results were generated retrospectively based on the full set of DFT coefficients as in [29]. Each subsampling method was evaluated in terms of SNR over a range of sampling ratios, considered the particular case $\gamma = 1/4$ for illustration. The fully sampled result of Section 4.2 was used as as a reference for evaluation.

Results of this experiment are shown in Fig. 3. Repetition subsampling (Fig. 3a) yielded the worst result due to noise. This noise issue is further exacerbated in the extreme case where the image is reconstructed from a single pass (Fig. 3e), which illustrates the need to use high

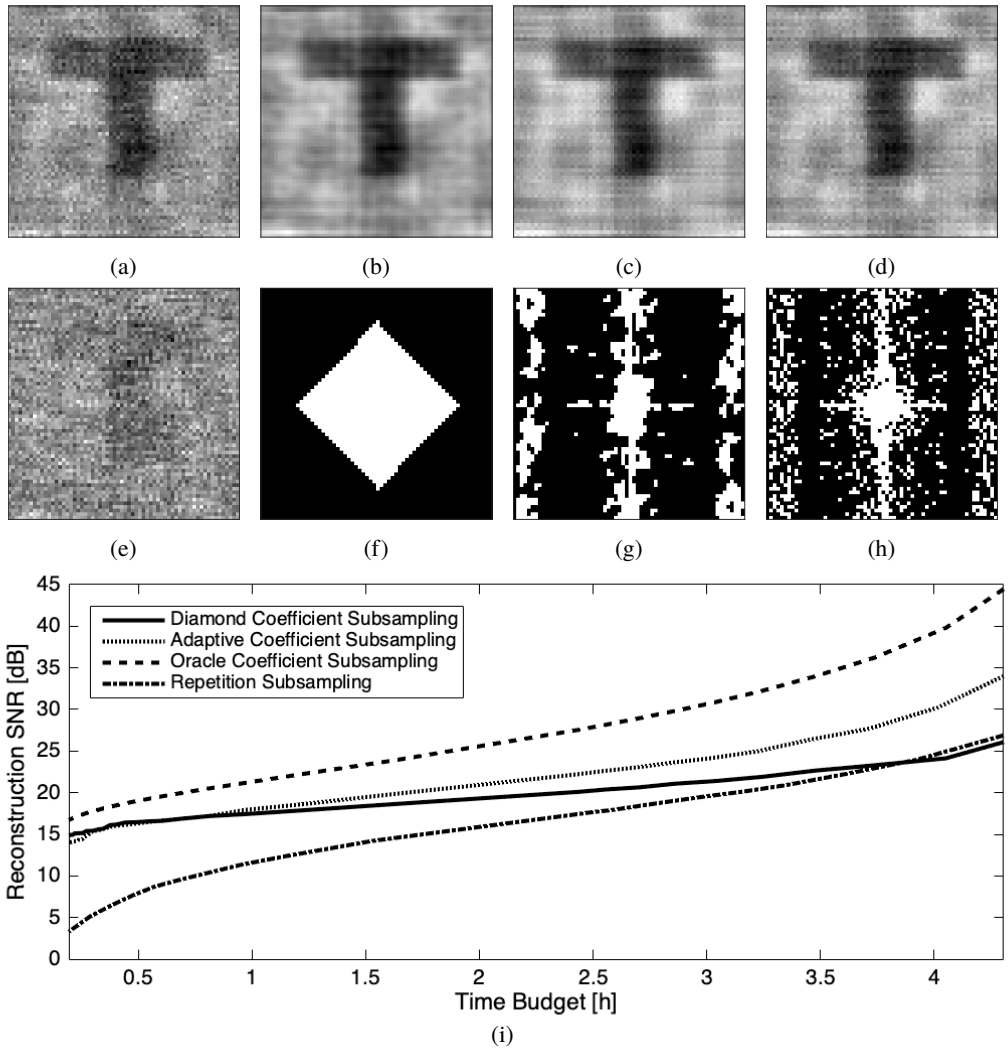


Fig. 3: Comparison between the different subsampling methods used in our experiment. (a)-(d) Reconstructions for each method under a limited time budget of 1.2 hours ($\gamma = 1/4$). From left to right, repetition subsampling (SNR = 12.37 dB), diamond coefficient subampling (SNR = 17.87 dB), adaptive coefficient subsampling (SNR = 18.59 dB), and oracle coefficient subsampling (SNR = 22.22 dB). (e) Reconstruction in the extreme case of a single pass ($\gamma = 1/50$) for repetition subsampling (SNR = -0.12 dB). (f)-(h) Acquired DFT coefficients (white pixels with redundant coefficients also shown) associated with the results (b)-(d) shown right above. (i) Reconstruction SNR in decibels for each method as a function of the allocated time budget in hours, varying γ accordingly.

enough values of L in our acquisition setup. Compared to both non-adaptive approaches (Figs. 3a and 3b), the proposed adaptive coefficient-subsampling method (Fig. 3c) yielded the best result. The latter was closest to the ideal oracle coefficient-subsampling solution (Fig. 3d) where the highest-energy coefficients are determined and selected a posteriori from the full acquisition ($L = 50$, $K = N/2 + 2$). This result was also less blurry than when using the non-adaptive

coefficient-subsampling method, which confirms the potential of adaptive approaches to better preserve high-frequency information. The spatial distribution of the coefficients selected by the adaptive method (Fig. 3g) was also close to the one of the oracle solution (Fig. 3h), as opposed to the non-adaptive case (Fig. 3f). The adaptive coefficient-subsampling method consistently outperformed its non-adaptive counterpart over a large range of sampling ratios (Fig. 3i). Its performance ultimately dropped at very low sampling ratios because it uses a constant time-budget overhead to determine the relevant coefficients. Overall, the above results highlight the critical impact of the subsampling strategy on the reconstruction quality, and the potential of adaptive methods to preserve reconstruction quality under smaller time budgets.

5. Conclusions

We proposed a novel imaging paradigm where an optical device can be employed in conjunction with a SLM to acquire and resolve remote objects with a resolution that exceeds its native diffraction limit. Our experiments represent the first proof of concept that the loss of spatial and angular resolution that is intrinsic to diffraction can be circumvented in this case, thanks to the use of a specific acquisition and reconstruction strategy. Our paradigm potentially extends to a relatively broad class of optical devices and applications. In that regard, it is important to note that the only property of $\mathcal{D}$ that is exploited to derive our result (29) in the Appendix is its preservation of the total signal energy. For this reason, (29) would also hold if the point-spread function in (2) were not isoplanatic, or if the light emitted from the object were spatially coherent, in which case diffraction effects would become non-linear in intensity but remain energy-preserving.

An important limiting factor that remains in 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. We showed how particular acquisition strategies can mitigate this issue to some extent. In further work, the proposed acquisition and reconstruction methods could be adapted to decrease the time budget, maximize the reconstruction quality, and operate in more complex settings. For instance, satisfactory reconstructions could be obtained from fewer coefficients based on the compressed-sensing framework [1, 2]. The latter is particularly well suited to our acquisition paradigm where image coefficients are measured in the Fourier domain. From a compressed-sensing perspective, our acquisition strategy can actually be seen as the transformation of the initial optical device that produces *degraded measurements*—due to diffraction and instrument response—into a new device that produces *compressed measurements*. To maximize the reconstruction quality when acquiring fewer coefficients or in the presence of noise, wavelet priors, total-variation priors, or priors based on non-local means could be used [30]. Adaptive-acquisition methods that are more advanced than the one proposed in Section 3.4 could also be developed. Such methods could 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 previous measurements. Furthemore, acquisition settings where the object is moving or changing could be potentially handled by adapting our acquisition model and our algorithms. This could be achieved using deep learning, as was already done for video acquisition with a single-pixel camera [31].

While the strength of the current work is its broad applicability to a large set of optical devices, dedicated optics could also be envisaged, drawing from previous works on one-pixel cameras, for instance. In current single-pixel cameras, one major difference is that the SLM is primarily meant to compensate for the lack of multiple sensors and 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, single-pixel architectures could

thus be adapted to implement our approach, and its single sensor could physically replace the numerical-integration process in (7).

In terms of applications, our paradigm could be employed to either better resolve objects, as was done in our experiments, or to better track them, considering adaptations similar to [31]. The proposed approach could also be exploited to image objects whose distance z_o is out of reach either for practical or for technological reasons, *e.g.*, at infinity. In that regime, the ability of our method to increase angular resolution, or, equivalently, to increase spatial resolution independently of the object distance, becomes key as there is not way the object can be accessed, illuminated, or approached to increase its perceived size in the image field of view. This is typically the case in astronomical imaging, in which case an additional SLM placed at sufficient distance may further enhance the resolution capabilities of an existing ground or space-based telescope. Such a configuration would bear similarities with the *external occulter* concept [32], except that light would be structured by a SLM instead of being blocked. For astronomical applications, the robustness of our paradigm to diffraction and the differential nature of our measurements may also prove useful to mitigate the effects of atmospheric seeing, depending on the relative location of the SLM. This remains the topic of further investigations.

6. Appendix

6.1. Unbiasedness of acquired modulated images

Under our noise model (14), we demonstrate below that the recomposed modulated images $\boldsymbol{g}_k^\delta$ defined in Eq. (15) are unbiased estimates of the modulated images $\boldsymbol{g}_k$ defined in Eq. (6).

First, the differential nature of (15) implies that the additive biases that are intrinsic to our noise model, which includes the effect of ambient light, cancel out in the expected value of $\boldsymbol{g}_k^\delta$. This, along with the fact that the $\boldsymbol{g}_{k,i,\ell}^\delta$ are noisy realizations of the same noiseless images $\boldsymbol{g}_{k,i}$, implies that

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

Combining the above with Eq. (13), we then obtain

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{g}_k^\delta\right] \quad = \quad & (\mathcal{SD}\{q_{k,1}(M\cdot)f(M\cdot)\} - \mathcal{SD}\{q_{k,2}(M\cdot)f(M\cdot)\}) && (22)\\ & +\mathrm{j}(\mathcal{SD}\{q_{k,3}(M\cdot)f(M\cdot)\} - \mathcal{SD}\{q_{k,4}(M\cdot)f(M\cdot)\}) && (23)\end{aligned}$$

Given the splitted structure of q_k in Eq. (12), and by virtue of the linearity of our forward model,

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

which, given the definition of $\boldsymbol{g}_k$ in (6), finally implies that

$$\mathbb{E}\left[\boldsymbol{g}_k^\delta\right] = \boldsymbol{g}_k, \tag{25}$$

which is what we wanted to show.

6.2. Invariance of estimated dot products to diffraction and sampling

Based on the result of Section 6.1, we demonstrate below that the estimates v_k^δ are robust to diffraction and sampling, and that they are equivalent to inner products between the degradation-free image f and the modulating patterns q_k, up to a multiplicative constant.

Given Eqns. (6), (7), and (24), the estimated dot products v_k^δ satisfy, for $1 \leq k \leq K$,

$$\begin{aligned}\mathbb{E}\left[v_k^\delta\right] &= \boldsymbol{u}^\top \boldsymbol{g}_k \\ &= \boldsymbol{u}^\top \mathcal{S}\mathcal{D}\{q_k(M\cdot)f(M\cdot)\}. \end{aligned} \tag{26}$$

Substituting $\mathcal{S}$ and $\mathcal{D}$ with their definitions in Eqns. (1) and (2), and given the definition of the unit vector $\boldsymbol{u}$, we obtain

$$\mathbb{E}\left[v_k^\delta\right] = \Delta_\mathrm{t} \sum_{p=1}^{P} \left[\phi * h * (q_k(M\cdot)f(M\cdot))\right](\mathbf{x}_p). \tag{27}$$

Since ϕ models the signal-integration effect of the sensor pixels on the sensor plane Σ_i, the above sum of samples corresponds to the integral

$$\mathbb{E}\left[v_k^\delta\right] = \Delta_\mathrm{t} \int_{\Sigma_\mathrm{i}} \left[h * (q_k(M\cdot)f(M\cdot))\right](\mathbf{x})\mathrm{d}\mathbf{x}. \tag{28}$$

Finally, since h is a diffraction operator, it is energy-preserving, and thus preserves the total light intensity that can be collected from the object. Assuming that the light stemming from the object is fully modulated by the SLM on its plane Σ, and then fully collected by the sensor array, in accordance with Fig. 1, the above equation simplifies to

$$\begin{aligned}\mathbb{E}\left[v_k^\delta\right] &= \Delta_\mathrm{t} M^{-2} \int_\Sigma q_k(\mathbf{x}) f(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \Delta_\mathrm{t} M^{-2} \left\langle q_k, f \right\rangle, \end{aligned} \tag{29}$$

which is independent of $\mathcal{D}$ and of $\mathcal{S}$, and which only involves the degradation-free image f.

6.3. Discretization

Each modulating pattern $q_k(\mathbf{x})$ is obtained from a complex-valued sequence $\boldsymbol{q}_k = [q_k^1, \ldots, q_k^N]^\top$ that determines all pixel values of the SLM array in each sub-acquisition. This relation is

$$q_k(\mathbf{x}) = \sum_p q_k^p b(\mathbf{x} - \mathbf{x}_p), \tag{30}$$

where $b(\mathbf{x})$ is a square box function of size Δ that represents the shape of the SLM pixels, and where q_k^p is the value of the p-th pixel of sequence $\boldsymbol{q}_k$. This makes Eq. (29) expand as

$$\begin{aligned}\mathbb{E}\left[v_k^\delta\right] &= \Delta_\mathrm{t} M^{-2} \int_\Sigma \sum_p q_k^p b(\mathbf{x} - \mathbf{x}_p) f(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \Delta_\mathrm{t} M^{-2} \sum_p q_k^p \left(\int_\Sigma f(\mathbf{x}) b(\mathbf{x} - \mathbf{x}_p)\mathrm{d}\mathbf{x} \right). \end{aligned} \tag{31}$$

We then define a discrete version of $f(\mathbf{x})$ as the sequence $\boldsymbol{f} = [f_1, \ldots, f_N]^\top$, with a discretization that exactly matches the resolution and pixel size of the SLM. Specifically, we have

$$f^p = \Delta_\mathrm{t} M^{-2} \int_\Sigma f(\mathbf{x}) b(\mathbf{x} - \mathbf{x}_p)\mathrm{d}\mathbf{x}. \tag{32}$$

For each measurement v_k^δ, this finally yields the simple discrete-scalar-product relation

$$\begin{aligned}\mathbb{E}\left[v_k^\delta\right] &= \sum_{p=1}^{N} q_k^p f^p \\ &= \boldsymbol{q}_k^\top \boldsymbol{f}. \end{aligned} \tag{33}$$

6.4. Adaptive subsampling

Subsampling methods such as the low-frequency diamond scheme proposed in [24] select coefficients a priori, regardless of image properties. They can thus miss relevant higher spatial frequencies. Inspired by [29], we propose an adaptive subsampling scheme that mitigates this problem. Exploiting our acquisition approach where L is typically greater than one, the key idea of our adaptive scheme is to first make one quick but exhaustive acquisition that already allows to estimate the *positions* of the highest-energy coefficients. Retaining K significant coefficients, and using L acquisitions per significant coefficient in total, it proceeds as follows:

1. Make preliminary estimates $v_k^{\delta,1}$ of all v_k using a single (instead of L) acquisition pass
2. Low-pass filter the 2D map associated with the DFT-coefficient estimates $v_k^{\delta,1}$ with a Gaussian filter of unit variance; this yields the filtered coefficients $v_k^{\delta,1\prime}$
3. Determine Ω as the set of positions of the K highest-energy filtered coefficients $v_k^{\delta,1\prime}$
4. Refine the first-pass estimates $v_k^{\delta,1}$ by acquiring coefficients $L-1$ more times if they belong to Ω—averaging results according to (15)—and replacing them by zero otherwise
5. Use the final estimates v_k^δ that are obtained for reconstruction

The low-pass filtering in Step 2 and the elimination of non-significant coefficients in Step 4 are used to reduce the effect of noise in the preliminary estimates $v_k^{\delta,1}$ that are obtained from a single acquisition pass. The time budget associated with the use of this adaptive scheme is

$$t_{\text{sub}} = (4KL + 2N + 8 - 4K)\Delta_{\text{t}}, \quad K < N/2 + 2, \tag{34}$$

the overhead $(2N + 8 - 4K)\Delta_{\text{t}}$ being associated with adaptive coefficient selection in Steps 1-3.