EXECUTIVE MASTER

STATE OF THE ART REPORT

Artificial Intelligence
The Executive Master is a unique diploma from École Polytechnique targeted to senior managers that combines technology, innovation management and leadership.

This part-time program is designed in modules and allows participants to follow the courses while respecting their professional commitments.

The modules cover various technological and innovation fields such as artificial intelligence, Data Science, biomimicry, IoT, new energies, sustainable development, industry 4.0 or biomedical engineering. The courses are given by researchers working in the laboratories of Ecole Polytechnique and of our partner institutions.

The knowledge acquired throughout the program helps participants to carry out a Team Project and a personal project, the “State of the Art” (SOTA) Report. The objective of the SOTA Report is to produce an updated synthesis on a scientific or technological subject. The document has to be quite short and readable by non-experts.

Over the years, many of these outputs have proven to be both intellectually very stimulating and practically very useful to accelerate the structuration of innovation processes\(^1\). Hence many readers suggested that we make them available to a larger audience.

This SOTA Report is part of a selection we did and for which the authors – now our Executive Master alumni – gave us their authorization to publish and share.

We hope you will enjoy exploring this “treasure”. Do not hesitate to send us some feedbacks or get in touch directly with the authors.

The École Polytechnique Executive Master team

---

\(^1\) Disclaimer: these reports are written under the sole responsibility of their author.
COMPUTATIONAL IMAGING

Key words:
Computational imaging, diffractive optics, DOE, metalens, deep-learning, lensless imaging, end-to-end optimization, aperture coding, Hyperspectral imaging, TeraHertz, super-resolution.

Executive summary:

Optic is a very active field of science, and so are computer sciences, image processing, material sciences in general and more specifically nanotechnologies. Combining the new developments in these areas led to a new (or at least renewed) paradigm called “computational imaging”.

The fundamental concept in computational imaging is to consider optics, electronics and computing algorithms as a whole and use a simultaneous design strategy for these three key components of an optical system.

The state-of-the-art developments in this co-design procedure rely on several technological enablers:

• The capability to produce at low cost optical components that enable to manipulate light in a very customized and flexible way (like focusing, phase shifting or changing polarization state...). This is based on several small-scale manufacturing technologies including the ability to use controlled nano-scale structures on optical gratings.
• The improvements in sensitivity, shape factor or cost of photo-sensitive devices like sensor arrays.
• The strong evolution in the last year of machine-learning algorithms with a specific focus on deep-learning applied to computer vision.
• The ability to solve one single numerical optimization problems that considers the whole system (optics and image processing), thanks to adequate formalism, high computing power and optical simulation capabilities.

In the first section of this document, the “computational imaging” paradigm will be detailed and illustrated with the problem of “aperture coding”, which is one of the very fruitful examples of joined design of optics and image processing. The second section will focus on the technological enablers that led to the strong development of the computational imaging approach. The recent advances in the field of diffractive optics, metalenses, spatial light modulators and deep-learning will be presented from the point of view of computational imaging. Finally, the third section will present recent successful applications of the computational imaging framework in many practical optical problems like 3D and spectral imaging or camera and microscope miniaturization. Finally, some perspectives and fundamental limitations to be overcome will be discussed in the conclusion.
# Table of contents

COMPUTATIONAL IMAGING .................................................................................................................... 1  
Executive summary: ............................................................................................................................. 1

INTRODUCTION ........................................................................................................................................... 3

1. PRINCIPLE OF COMPUTATION IMAGING .................................................................................... 4
   a. What is an optical system? .............................................................................................................. 4
   b. Optical design and optimization ................................................................................................. 6
   c. The simple example of aperture coding .................................................................................... 7
   d. Conclusion and synthesis ............................................................................................................ 9

2. TECHNOLOGICAL ENABLERS .................................................................................................... 9
   a. Manufacturing dedicated optics .................................................................................................. 9
   b. Diffractive optical elements and metalenses ............................................................................ 11
   c. SLM ........................................................................................................................................... 13
   d. Optical simulation ...................................................................................................................... 14
   e. Deep learning ............................................................................................................................ 14
   f. Conclusion and synthesis ............................................................................................................ 15

3. FIELDS OF APPLICATION ........................................................................................................... 16
   a. Compressive sensing and super-resolution ............................................................................... 16
   b. Lensless imaging and small shape factor .................................................................................. 17
   c. 3D imaging ............................................................................................................................... 18
   d. Phase imaging ........................................................................................................................... 19
   e. Spectral imaging ....................................................................................................................... 19
   f. THz imaging .............................................................................................................................. 21
   g. Conclusion and synthesis ............................................................................................................ 22

4. CONCLUSION - LIMITATIONS AND PERSPECTIVES .................................................................. 23
   CONCLUSION ................................................................................................................................. 23
   LIMITATIONS ................................................................................................................................. 23
   PERSPECTIVES .............................................................................................................................. 23

APPENDIX 1 – Glossary .......................................................................................................................... 24

APPENDIX 2 – PSF computation from DOE height field ...................................................................... 25

ACKNOWLEDGEMENT AND READING ADVICES: .......................................................................... 26

SOURCES ............................................................................................................................................... 27
INTRODUCTION

Capturing images is a very old discipline of art and science whose applications in our day to day life are tremendous. The history of imaging goes back in the Antiquity when the first optical systems were designed and reported. Key milestones have progressively been reached with the understanding of many light related phenomena, like Snell’s and Descartes’ law of refraction or Fraunhofer work on spectral scattering. The ability of recording images in the first part of the 18th century is a huge step forward in many areas. As maybe an emblematic success of the new born photography discipline, Eadweard Muybridge was able to understand the complex mechanics of galloping horses illustrating already the great potential of imaging technology for studying fast phenomena. But of course, this was just the beginning of the story as more complex optics theories were built, like Maxwell’s equations in 1865 and a better understanding of the optical properties of specific material like glass was also progressively acquired. With both theory and practical material knowledge, it became possible to “engineer” more sophisticated instruments for different applications like microscopy. The last revolution in imaging has probably been the emergence and then very fast spreading of digital recording and digital image processing with the invention of the CCD detector in 1969 which opened the door to the automatic interpretation of image data with computers and processors.

For a long time, the principle of imaging systems has remained quite similar. The key idea was to design optics to collect light that would keep artifacts, like chromatic or spectral aberrations, at a minimum level and to combine this system with a high-quality acquisition device, like a ccd or cmos matrix keeping also the “electronic noise” at a minimum level to obtain an image that can be related as closely as possible to the “real-world” scene imaged through the device. The image obtained through any kind of digital recording could then be used as an input for enhancement algorithms, like denoising, geometric distortion correction and/or automatic interpretation algorithms like identifying the objects captured in the image.

The process of designing an automatic image processing pipeline was usually a sequential process:

- Designing the best possible acquisition device for a specific application, usually considering a trade-off between performances, cost and specific constraints (like components availability, weights, size, robustness…).
- Designing the best possible algorithm to process the images acquired by the device.
- Integrating the whole system into a complete acquisition/processing and possibly actuation chain if the output of the processing is used to trigger automated reaction.

Of course, it is rarely completely sequential, and there are many interdependencies between the design of the acquisition device and the image interpretation process. But despite this, in most of the situations, the work on data processing starts when the acquisition device is operational and data sets are available for testing the algorithms.

The images obtained by an acquisition system are usually self-understandable and their quality is quantified by how accurately they represent the “real” object under study, but it has been an important tendency in many areas to use computer codes as a key step to transform an otherwise ciphered information into interpretable data. Common examples are well-known in medical imaging: X-Ray tomography or magnetic resonance imaging require a complex step of raw data pre-processing involving dedicated algorithms, like the Radon transform, to “compute” a 3D representation of the organ or body part of interest.

The core concept of “Computational Imaging”[1] can be defined as relying on dedicated algorithms or computation steps to retrieve information about an object of interest that could not be accessed through a direct measurement.

In the past years, this way of approaching imaging problems has gain a strong interest maybe due to the increasing computing power available at low cost, making it less and less impacting in terms of price to add computing resources to imaging systems, even in individual devices such as smartphone.

As an illustration of the interest for the field, the IEEE launched a journal, IEEE transactions on Computational Imaging in 2015.

The target of this report is to present a state of the art of the main concepts in Computational Imaging as well as the main impacting applications in fields such as personal imaging, industry or microscopy. In the first section, the principle of computational imaging is described in more details based on classical examples. In the second section, it will be shown, how a set of “technological enablers” such as the
evolution and miniaturizing of specific optical components or the development of optical modeling framework or deep-learning technologies are contributing to the extension of the field of applications of computational imaging. In the third section several concrete and developing examples of applications will be covered from spectral to 3D imaging. Finally, some market and technological perspective related to those technologies will be shortly discussed in the conclusion.

This document is mostly a review of existing literature with no original scientific work. The choice has been made to cover many examples and many different fields related to computational imaging from hardware to algorithms: many references (most often freely available on the internet) are provided in the text so the interested reader can find the details which could not be included in this document.

1. PRINCIPLE OF COMPUTATION IMAGING

a. What is an optical system?
An optical system is the set of components leading to the formation of an image. The details can of course depend on each use-case, but Figure 1 below illustrates the most common situation:

![Figure 1: most common components of an optical system](image)

In Figure 1, the following components are considered:

- **One or more light sources** $S^n_{\Psi}(S_n, \hat{n}, t)$.
  $S^n$ represents the light intensity emitted from point $S_n$ in direction $\hat{n}$ at time $t$ by the $n$th light source of the system. $\Psi$ stands for all parameters considered in the problem over which it is interesting to marginalize the information. The most common quantities would be the wavelength $\lambda$ (sometimes simplified as a color channel) or the polarization state. Several types of light sources can of course be considered (laser, LEDs, natural lights...) depending on the problem under study.

- **An object or scene** to be imaged:
  It can be described with different formalisms, like the 3D shape and the bidirectional reflectance distribution function (brdf), which quantifies how much energy is absorbed or reflected depending on its incidence direction relative to the surface normal. In more complex situations, some other dependencies or phenomena can be introduced in the object description like complex refractive indexes for transparent media. In summary, the precise nature of the parameterization of the scene depends a lot on the mechanisms and approximations that are considered in the image formation.

- **The collection optics** is the core object that will be discussed in this document. In the most common
optical systems, it is made of one lens (or a set of lenses) whose function is to focus the light in one focal plan where the photo-sensitive detector is located. Much more complex optical devices will however be discussed in the following sections. The collection optics can be parameterized by a transfer function $\Delta_{uv}(M, x, y, t)$, which will transform the light wave coming from point $M$ by changing its intensity and adding a phase shift which can depends on the position $(x, y)$ on the collection plan, and also on time for active systems that can be modified on demand (see section 2.c). The dependency of the transfer function to the wavelength or polarization state can be expressed (if needed) through the parameters $\Psi$, and $v$ stands for all degrees of freedom that can be used to tune the behavior of the optical system. In the case of a classic lens, it could be the lens curvature and refractive index. Usually, the transfer function is not used directly in the computation that we will discuss later, but it is used to compute a very important characteristic of the optical system which is the Point Spread Function (PSF). This function $P_{uv}(M, u, v, t)$ characterizes how the light from a single point $M$ is spread on the focal plan of the detector array (Figure 1). If the optical collection is a perfect lens and $M$ is in the object focal plan, then the PSF is a Dirac function located at the conjugate position in the image focal plan on the other side of the lens. But if the point $M$ is not in the focal plan, every photograph knows that a “blur” effect will be observed in the picture, so the intensity coming from point $M$ is spread over several pixels of the image. The effect is obviously increasing with the distance to the focal plan, which can be strong asset to evaluate the depth of the object (section 3.c). The PSF is much easier to manipulate in computation than the transfer function $\Delta$ to simulate an acquisition through the optical collection system, but the use of $\Delta$ is often required when the design of the optical system must be optimized thanks to the degrees of freedom represented by $v$. This would help answering questions like: “what kind of lens curvature should I used for imaging a scene in a specific geometric configuration?”.

- **The sensor array** is the imaging device strictly speaking. It is most often based on photo-sensitive electronics like semi-conductor cells. CCD and CMOS are the most commonly used for the visible wavelength range between 400 and 800 nanometers, but other semi-conductors or even other technologies like bolometers (temperature sensitive devices that heat with the amount of collected light) for infrared. This device is usually a rectangular array of cells that produces an image $i(u, v, t)$, $u$ and $v$ being the coordinate system in the resulting image. However, it can happen that more complex geometries (like curved detectors) are considered, and it of course does not change the principles of computational imaging that will be presented below: only the complexity of the modeling can be increased. On the opposite, some situations might be simpler, like in the case of the “single-pixel” camera described in the section related to Terahertz imaging (section 3.e), where there is not matrix of sensor but a single photo-sensitive cell.

- **An algorithm or computing procedure $\Phi_{\omega}$** is used to transform the image (or a set of images) into a “final” result $r_t = \Phi_{\omega}(i(u, v, \tau), \tau < t)$.

  Image processing has now become a very widely used and common field of computer science. Many libraries, algorithms and references are easily available on the internet, and it would be illusory and out of scope to make a comprehensive review of image processing technologies in this document. However, in the field of computational imaging, computation is of course a key step. The computing procedure $\Phi_{\omega}$ can take many different expressions like explicit analytic functions of the pixel values of an image, the result of several iterations of a simple transformation procedure, the comparison with reference data base or abacus… Finally, $\Phi$ should be really considered as a computer code parameterized by a set of values $\omega$ that can be used to tune its behavior. The result $r_t$ obtained after applying the computing procedure $\Phi$ to the images $i(u, v, \tau)$ captured by the sensor array can be another image (or set of images) corrected from various artefacts induced by the collection optics. It can also be a measurement, or a set of features computed from the image data, like the dimension of an object, its radiance, its position in the 3D space. In the case of facial recognition, the final result is simply a name. The fact that the output of the optical system is not necessarily an image of the scene is one of the foundations of computational imaging, which can lead to different optimization in system design than improving the intrinsic quality of the image $i(u, v, t)$. 


b. Optical design and optimization

The overall topic of computational imaging can be synthesized in the following question: **How to design the optical system to obtain the best performances at the exit of the computing layer in terms of retrieving the features of interest, while still fulfilling certain constraints?**

The design constraints are linked to practical degrees of freedom in the light source, collection system or imaging device like the form factor, weight or price of the system.

Consider a photographer expecting to make the best picture of the landscape in front of him. The degrees of freedom at its disposal are relatively limited: he can select one of the few objectives in his handbag and select a few parameters in its reflex settings (sensitivity, exposure time, aperture). His “cost function” for optimizing the system can also be defined in several ways: best fidelity regarding the sunset colors, best rendering of the small flower in front of him or capturing a fast movement in a sport scene without blur. Most of the time, all the target cannot be reached with the same settings... In that case, the problem can be expressed in a simply way, and the photograph must rely on his experience to make the choice.

For more complex situations, it is useful or even necessary to model the optical system either through analytic or numerical expressions of the output corresponding to a typical input scene. This enables to cover by means of simulation the range of possible settings for optimizing the output according to objective criteria. Many examples will be provided in the following sections but optimizing the geometry of a lens to limit chromatic aberrations, selecting the power and color of a light source are very common examples. In classical situations, the intrinsic quality of the resulting image (sharpness, depth of focus, noise level...) is often used as an objective criterion for system optimization and “translated” into mathematical forms by a cost function $\Lambda$.

From a more formal point of view, this is equivalent to select the set of parameters $\nu$ of the optical collection system by optimizing the cost function $\Lambda$ which depends on the transfer function $\Delta_{\nu}$ and usually a set $S^1 = \{(o_q), q \in [1, N]\}$ of typical objects to be imaged:

$$\hat{\nu} = \text{argmin}_{\nu}(\Lambda(\Delta_{\nu}, S^1))$$

Once the best optical design $\hat{\nu}$ has been found and manufactured, it becomes possible to acquire data and images with the system and obtain a set $S^2 = \{(i_k, r_k), k \in [1, N]\}$ of image examples associated with the expected results for each of them. Then a new cost function $\Gamma$ can be drafted to optimize the algorithm $\Phi_{\omega}$. $\Gamma$ can take many different expressions, but the most common is based on the sum of differences between the computed result using $\Phi_{\omega}(i_k)$ and the ground truth result $r_k$:

$$\Gamma(\omega, S^2) = \sum_k ||\Phi_{\omega}(i_k) - r_k||^2$$

The best set of parameters of the algorithms is therefore obtained by minimizing $\Gamma$:

$$\hat{\omega} = \text{argmin}_{\omega}(\Gamma(\omega, S^2))$$

Doing so, there is no coming back to the optical design when working on the algorithm and the optimization is fully sequential: the collection optics is optimized first, and the algorithm is defined and optimized in a second step. However, all optimization theories state that the conjunction of two local optima $\hat{\nu}$ on the one hand and $\hat{\omega}$ on the other hand only rarely realize the real optimum of the global system which consists of the composition of $\Delta$ followed by $\Phi$ for a typical scene containing an object $o$:

$$r(o) = \Phi_{\omega}(\Delta_{\nu}(o))$$

If the expected results $r_q$ are associated to the objects $o_q$ contained in the data set $S^1$, it becomes possible to define the new data sets:

$$S^3_{\nu} = \{(\Delta_{\nu}(o_q), r_q), q \in [1, N]\}$$
The joined optimization problem sometimes called “end-to-end” optimization of the couple \((v, \omega)\) consists in solving the following problem:

\[
(v^*, \omega^*) = \arg\min_{(v, \omega)} \left( \Gamma(\omega, S^2_v) \right)
\]

Even if in many situations, the successive optimization of the components of the optical system already provide good practical results, considering the global problem of joined optimization of the optical collection system and algorithm tuning together has been one of the important evolutions of computational imaging in the last years. This tendency is driven by the evolution of simulation and optimization tools and opens the door to many new imaging concepts that will be described in the following sections.

In [2], Stork and Robinson provide a quite generic optical modeling framework to conduct such a joined optimization. They used the Zemax optical simulation software and a Wiener filtering framework as main image processing code in the joined optimization loop to reach a global optimum which provide less sharp raw images from the camera (due to “sub-optimal” optics in terms of signal to noise ratio), but which combined with specifically tuned Wiener filters lead to a better image after reconstruction.

In Figure 2, the left column corresponds to the result of the sequential optimization problem: \(\hat{v}\) leads to the top left image which is the sharper image that can be obtained with the low-cost, low resolution optical device considered in the study. However, the deblurring algorithm applied on this image cannot get rid of some artefacts. The right column is the result of the joined optimization problem leading to a much better result with less artifacts, despite the visually lower quality of the raw data acquired by the sensor array (top right image).

This might be considered as a marginal improvement and this is indeed a toy problem to make the concept of joined optimization clear, but significant optical system improvements will be demonstrated in the following sections in many practical situations.

c. The simple example of aperture coding

Aperture coding is a concept that enabled a lot of successes in many areas like X-ray imaging or 3D imaging with a single camera [3]. If it is not exactly a “state of the art” technology by itself, it is a very rich framework, very illustrative of the computational imaging principles and still widely applied with new optical components and new machine learning methods. Several other applications will be described in section 3.
The concept of the pinhole camera is very simple (see Figure 3): if a detector array is placed in a box in front of a tiny hole, only the light coming from a specific direction will hit the detector at a specific position. The smallest the size of the hole (called the aperture), the sharper the image will be, and if the hole is enlarged, then the image become blurred since light originating from different close locations will hit the detector plan at the same position. This specific method of imaging is an alternative to designing optics with lenses to focus light, and it has been used a lot for wavelengths range such as X-rays where “optics” are hard and/or expensive to design.

Unfortunately, while image sharpness will require small pinhole (until diffraction limit is reached), the signal to noise ratio will on the contrary increase with the area of the pinhole since more light will globally enter the detection system with a bigger aperture. A trade-off must then be found between image sharpness and noise level. To increase the signal to noise ratio, more light must enter the dark box, and if the aperture must remain small, a natural idea is therefore to use many small holes instead of a single big one.

Very intuitively, shifting the hole from its original position is like shifting the detector array, and the effect of having multiple holes will be like superposing several shifted images of the same scene (Figure 4). A computation will therefore be needed to re-compute the original image knowing the position of the many holes.

Let consider a very simple modeling of the process of imaging an object $O$ with an aperture mask $A$ presenting a set of holes. In a rough approximation, the obtained image $I$ is simply the convolution of the mask with the original object:

$$ I = O * A + N $$

$N$ is a random noise term.

This is a very bare approximation, and in many applications more complex modeling can be used like using phase factors to consider also diffraction phenomena when the size of the holes becomes of the same order of magnitude than the wavelength of the light. But even in this simple expression, the formalism can lead to interesting developments.

An estimate of the original object can be obtained by a using a well-chosen “reconstruction” kernel $G$:

$$ \hat{O} = I * G = O * (A * G) + N * G $$

If the pair $(A, G)$ is designed so that $A * G$ is “close to” the identity matrix, then $\hat{O}$ can be a “good estimate” of the original object, since the noise level $N$ is low compared to the light coming from $O$ (because the global amount of light in the system is quite big thanks to a large number of holes). The “computational imaging” procedure is simply the process of optimizing simultaneously the pattern of the aperture mask $A$ (that is the parameter $\nu$ of the collection optics $\Delta\nu$), which can have specific constraints (number of holes, size of holes, total area of the mask…) and the reconstruction kernel $G$.
(that is, the set of parameters $\omega$ of the computing procedure $\Phi_\omega$). The optimal pair $(A, G)$ can also depend on the global optical configuration of the system, like the distance to the object and its expected size and/or shape if they are known. One of the classical optimization procedures leads to specific mask patterns called “uniformly redundant arrays” (URA)[4]. The exact same principle of “coded-aperture” masks can be used in front of a classical camera to help recovering 3D-information by analyzing the level of blurring in the image (section 3.a, 3.c).

d. Conclusion and synthesis
For complex imaging operations, an “optical system” in its globality must be considered. The system is usually made of optical components (light sources, collection optics, lenses, masks…), a set of light detectors (CMOS or CCD arrays, photo-diodes, bolometers…) and a computation algorithm to retrieve useful information from the raw data measured by the detectors. In classic systems, the optical part is designed and optimized independently from the image processing procedure, but in complex situations, when the number of degrees of freedom in the optical system becomes high or when the intrinsic quality of the optical system is hard to define without the reconstruction algorithms, it can be much more efficient to use a joined optimization of both optical components and algorithm parameters to increase the performance of the global system.

2. TECHNOLOGICAL ENABLERS
In this part, several components and technologies that act as key features in many modern optical systems will be reviewed. They are at the core of computational imaging, and the degrees of freedom they provide enable their flexible integration in many different optical settings for many different purposes like 3D or phase imaging, contrast enhancement, miniaturization that will be presented in section 3.

a. Manufacturing dedicated optics
One of the strong tendencies in the last decades and years has been the capability to design optical components with mastered shape and material at various scale, including very small ones, like micrometers and even nanometers (Figure 5) for surface patterning.

![Figure 5: Typical order of magnitude of the size of common objects (from [5])](image)

MEMS (micro-electromechanical systems) are an example of small scale systems involving electronics functions and mechanical components at micrometer scale in a single integrated device [5]. They are usually based on a silicon substrate over which several layers of polymers, oxides, gold can be grown or etched to reach the final design (Figure 6).
MOEMS (micro-opto-electro-mechanical systems) are MEMS involving small optical components that can be connected to miniaturized actuators. Typical examples are matrices of micro-mirrors that can be independently moved in various positions, leading to many applications in telecommunication, like optical switches, optical fibers connection (Figure 7), laser diode focusing [5]...

Micro-lenses arrays are a very representative example of optical design at micron scale. It consists of a usually flat area covered with adjacent lenses that can be used to focus light into several spots like optical fiber entries or specific detectors like CCDs, CMOS or photodiodes (Figure 8).

Typical state of the art involves hundreds of lenses over few millimeters. Application are numerous in plenoptics and 3D imaging (see section 3).

Another commonly used method for preparing thin optical components is photolithography[7]. The principle is to grow some layers on top of a substrate wafer, and to locally etch it, usually by chemical attacks. By etching at specific locations, it is possible to master the topography of the surface leading to specific optical behavior (see section 2.b). One of the most common process to select the area of etching is to use a “photoresist” product deposited by spin coating that can be activated by UV light exposure. After washing, it is possible to start the etching which will only happen at the locations not protected by the photo-resist (Figure 9). There are many variations around that process that can also be adapted to multilayer systems, but spatial resolution that can be achieved is around 1 micron whereas
the depth of etching can be around hundreds of nanometers.

Figure 9: Principle of photolithography (from [8])

High quality optical components with flexible shapes (Figure 10) can be also manufactured with subtracting manufacturing methods like machining, ultra-precision cutting, grindings or other technique[9]. It can be used to design small Fresnel lenses (Figure 11, Figure 11) or to obtain even more complex optical behavior by patterning the surface at scale smaller than the wavelength of the light that is intended to be used to obtain diffraction phenomena. Such methods enable to manufacture both discontinuous (with step-like structures) and continuous surfaces with high optical efficiency. Precision reached in the surface design can be around 30nm with very good polish (less than 2nm in roughness).

Figure 10: Example of free form optics manufactured using high-precision machining. From [9]

In the last years additive manufacturing technique[10] with methods such as selective laser melting (SLM), fused deposition modeling (FDM), stereolithography (SLA) or multiphoton stereolithography (MPS) which can achieve features of nanometer resolution over objects that have a global size in the order of magnitude of the millimeter[11]. These techniques are still at an early stage of development, but they are very promising to continue improving the degrees of freedom and optical quality that can be obtained in designing optics, those, providing more and more opportunities for computational imaging to enlarge its scope of applications.

b. Diffractive optical elements and metalenses

The origin of diffractive optical elements (DOE) was the idea of Fresnel lenses: replace a thick refractive optics with thinner systems realizing the same “light focusing” optical behavior:

Figure 11: Principle of Fresnel lens: the phase shift (and therefore the focusing behavior) from the left refractive optics is the same as the thin structure on the right. (from [5])

By shaping the surface of a thin transparent material, it is possible to realize complex phase functions. The possibilities are in fact much larger than replicating the behavior of a thick lens. The phase shift of
the light crossing the DOE at a specific position is proportional to the local thickness of the grating (Figure 12). By playing locally with the thickness of the material, that is, by mastering the surface topography (the height profile) with the right precision (~tens of nanometer), it is possibly to master accurately the spatial phase shift leading to a very customized optical behavior (see APPENDIX 2 – PSF computation from DOE height field). Typical order of magnitude obtained with classic photolithography for spatial resolution is the micrometer, and few hundreds of micrometers for the total thickness with steps of hundreds of nanometers. More precise surfaces can be obtained with other techniques (see previous section).

It is also possible to refine the DOE behavior using multilayer materials with different refractive index, which can help limiting chromatic aberrations for instance. Example of applications will be provided in the following sections.

As presented in the previous section, several manufacturing methods (small scale machining, 3D printing, lithography…) can be used with different results in terms of surface quality, precision, resolution. The research and development in these fields is very active.

With the same ideas in mind of controlling the phase behavior at the exit of a flat and thin optical components, the concept of diffractive elements has been recently refined with the concept of “metalens”[12]. In addition of having micrometer scale structure at the surface of an optical compound, some nanometer scale structures (below the targeted wavelength range) and sometimes called “nano-antennas” are also printed at the surface of the material (Figure 13, bottom right in particular), adding another level of control to the light behavior. A good review of the performances of such systems in terms of numerical aperture (the capability to focus light at a short distance of the optical system), achromatism (the capability to have good quality optics for a large range of wavelengths) and aberrations (capability of having a good optical quality without distortion over a large spatial dimension) can be found in [13]. Being a relatively new technology and a very active field of research in optics and material science, there are still some debates about the relative advantages of metalenses over classic diffractive optical elements presented above. See for instance [14] which emphasises in particular the capability of such system to better control light polarization or [15], in which authors stresses that better optical quality can be obtained with multilayer DoE for many applications (when polarization analysis is not at stake) but the sub-wavelengths dimensions of metalenses can be very useful in integrated photonics devices (see [16]).
Figure 13: Examples of metalens (from [12]) with associated legend: “Dielectric polarization-dependent flat lens. (A) Bottom: optical microscope image of a reflective lens consisting of an array of amorphous silicon groves with spatially varying width. Insets show scanning electron microscope (SEM) images of various locations of the lens. The lens is designed in the near-infrared (NIR) and has a numerical aperture (NA) of ~0.01. Top: measured beam profile in the focal plane of a reflective lens. Inset shows the measured beam radius (1/e²) along the propagation direction. (B) Bottom: SEM image of the center portion of the fabricated lens. Inset shows the measured focal spot intensity profile. Top: measured intensity distribution along the propagation direction showing the evolution of the beam before and after the focal spot. (C) Bottom: SEM image of the fabricated lens. This monochromatic metalens operates based on the Pancharatnam-Berry (PB) phase using titanium dioxide nanofins and it can be designed across the visible spectrum with high efficiency. For instance, three lenses with NA = 0.8 and diffraction-limited focusing at wavelength of 405, 532, and 660 nm were reported with corresponding efficiencies of 86, 73, and 66%. Top: measured intensity distribution in dB along the propagation direction.”

c. SLM
Spatial light modulators (SLM) are also devices that enable to shape the phase, intensity or polarization of an incident light beam, similarly to DOE, but in a controllable fashion. With SLMs, local phase shift can be changed on demand by electrical command contrary to DOE, which are basically “static” components. Typical SLM devices can have up to 1920x1200 pixels with pixel size close to 4 μm and a typical frequency of 60 Hz [17]. These devices can be used for generating holograms, adaptive optics (for correcting non-uniformity in atmosphere aberrations for instance) or shaping laser beams… The most common type of SLM uses liquid crystals that can be aligned at the pixel level under the influence of an electric fields (Figure 14). Aligning the crystals changes the optical properties of the pixel and those the phase of the light crossing the pixel or being reflected by it.

Figure 14: principle of spatial light modulators (SLM) (from [17])

Compared to DOE or metalenses, the phase control is not as flexible, and the resolution might be slightly lower but the capability of having a real-time control over the phase of the incident light opens a lot of potential applications. In additions, a lot of researches try to combine the metalens concept of nanostructures with the key principle of SLM based on the orientation of liquid crystals [18].
**d. Optical simulation**

As seen in the previous sections, classical refractive optics (lenses), diffractive optics (DOE and metalenses) or SLM, potentially combined with other optical devices can modify very accurately the properties of an incident light beam. But such many degrees of freedom make it of course difficult to integrate in global systems without accurate numerical simulation tools. For computational imaging, the expected output of the global optical system, like the capability of computing specific features related to the measured object or scenes, is the starting point of the design. It is therefore required to solve the inverse problem, which is the computation of how the optical devices should be engineered to achieve specific functionalities like numerical aperture, focusing capability, spectral scattering, polarization control or any other optical behavior.

The direct problem is already a challenging task when it comes to simulating the interaction between light and nanostructures, and the inverse problem is even more complex, but both are a key aspect of computational imaging.

Usually, the optimization related to the light collection system is conducted in two steps. In the first step, the global behavior of the optical components is defined through a first loop of optimization leading to the expression of an “ideal” Point Spread Function (PSF) for the optical system. In a second step of optimization, the best optical design in terms of geometric configuration of the optical components (number of layers, thickness of each layers, nanostructures size density and positions...) is computed for manufacturing the physical system realizing the “optimal” PSF.

As optical components like DOE, metalenses or sLM are developing, numerical capabilities to simulate their behavior are also progressing in all fields. For example, the PhD thesis of [8] proposed new models, algorithms and rapid prototyping techniques for DOE, whereas inverse problems for optimizing metalenses are also strongly investigated [19], [20].

Providing such simulation tools is also the object of several commercial companies like Zemax, Ansys and its Lumerical module or LightTrans international.

**e. Deep learning**

Deep-learning has been one of the strongly growing artificial intelligence framework in the last years. It led to many successes in the field of image processing where deep-learning based algorithms achieve state of the art performances in most classification and segmentation benchmarks, but also in the field of speech recognition, automatic document analysis and supervised learning in general.

Classification is a common problem where different input vectors must be divided in different categories, like deciding if an image contains a cat, a dog or something else (three categories in that toy example). There are different possibilities to solve such a problem. The first one is to define explicit criteria or rules that enable to decide about the category: if a vehicle has only 2 wheels and no motor then it’s a bicycle. However, in some situations, it is hard to draft by hand all rules that should be used to decide the category of an object, especially if the object is a high-dimensional vector such as an image. In that situation, one can use a “training” procedure based on examples of objects of different categories \( \mathcal{S} = \{(i_k, c_k), k \in [1, N]\} \), where \( i_k \) is an input vector, \( c_k \) is the associated category and \( N \) is the number of examples in the data set. For deep-learning, \( N \) is usually a large number (hundreds, thousands or millions but it depends a lot on the underlying distribution of input vectors to decide what “large” means). The training procedure consists in defining an approximation function \( f_\mu(i) \) that takes input vectors as argument and provide a category as an output. The set of parameters of the approximation function \( \mu \) is chosen so that the number of misclassified samples on a test set is minimized. Most commonly, this follow an iterative gradient descent procedure to reach an optimum choice for the vector \( \mu \). In the case of deep-learning \( f_\mu \) as a quite specific mathematical formulation as a composition of successive layers of “neurons”, which are simply linear combinations of relatively simple non-linear functions. In that case, the set of parameters \( \mu \) is called the set of “weights” of the neural network. Figure 15 illustrates some of the most common components of deep-learning architecture.
Figure 15: Classical deep-learning components. a) A labeled data set containing a lot of images with associated information (here a category from the imageNet data base) used for training the system. b) An elementary network structure based on linear combinations of simple non-linear functions. c) A convolutional neural network (CNN) to work with images, which is a customized architecture made of many iterative layers of one or more elementary structures.

In the case of images, this class of algorithms is matching very well with the concept of PSF (Point spread function), which can be used as a convolution kernel to compute the image captured by the detector array starting from the light field before the collection optics. It is sometimes straight-forward to express the global transfer function \( r(\omega) = \Phi_\omega(\Delta_\nu(\omega)) \) (section 1.b) from the light field entering the collection optics to the final result after the computing algorithms as a single differentiable function including PSF computation from height field \( \nu \) and a large neural network with weights \( \omega \) (Figure 16), those taking profit of all developments and progresses made in the field of numerical optimization and deep-learning in the last years. [21] presents a good synthesis of recent example of use of deep-learning within computational imaging framework.

Figure 16: End-to-end optimization (from [21]): Optical settings leading to system PSF and deep-learning based image processing can be included in one single differentiable model, that can be used, for instance to optimize object classification.

f. Conclusion and synthesis

In this section, we have reviewed several optical components (diffractive optical elements, metalenses) which can be used to modify in almost any fashion an input light beam. Some of these components (Spatial Light Modulators) can even be manipulated in real time to make the optical system flexible and adaptive. These components can be manufactured at relatively low cost, leading to a very wide space of possible optical design \( \Delta_\nu \). The images recorded by the sensor are further processed by any kind of algorithms \( \Phi_\omega \) (typically deep-learning). Thanks to numerical simulation, the whole process can be jointly optimized to retrieve the optimal features of interest from the scene or object.
3. FIELDS OF APPLICATION

The methodology presented above is very generic and can be applied in many different contexts, at many different wavelengths for very different purposes, leading sometimes to impressive results compared to traditional imaging.

a. Compressive sensing and super-resolution

Compressive sensing and super-resolution are classic approaches in image processing. The principle is to image a scene with a low-resolution sensor array (meaning a small number of pixels) and recover a high-resolution image through a computational strategy. The counter-part is that specific assumptions about the observed data have to be made (like some kind of sparsity properties in the Fourier or wavelet domain [22], [23], which is well verified for natural images). This usually involves aperture coding with random masks and a computation phase to retrieve the high-resolution image as the solution of an optimization problem.

SPAD (Single Photon Avalanche Diodes) cameras are imaging device with very high sensitivity and very high frame-rate (it can reach more than 10000 frames/second with low readout noise [24]). Despite its advantages, the electronics is more complex than in classical CCD or CMOS cameras, leading to lower resolution sensor arrays and lower fill-factor (the ratio between photon sensitive surface over the total surface of the printed circuit board). However, by designing the collection optics and using deep-learning based super-resolution techniques within a computational imaging framework (Figure 17), it is possible to retrieve a high-resolution (Figure 18) and even depth images with a low-resolution SPAD camera [25].

![Figure 17: Computational imaging optimization procedure applied in [25]. The PSF is tuned together with a super-resolution deep convolutional network in a first optimization loop using forward image simulation. Then the phase plate is build using DOE.](image)

The computational imaging framework enabled to multiply by 4 the resolution (Figure 19) with accurate results (no aliasing and low noise) opening interesting perspectives for this camera.

![Figure 18: Results obtained with a fast SPAD camera combined with optimized optical phase mask [25]. i) Low-resolution raw data of the sensor. ii) Super-resolution results: the high frame rate (1250 image/s) enables to follow accurately the movement of the fan. iii) reference image with classical high-resolution camera.](image)
b. Lensless imaging and small shape factor

One of the great advantages of diffractive optics is their small form factor and low price compared to high quality lens systems. “Lensless” imaging finds two very important applications for designing very small and lightweight cameras and for digital microscopy where high-magnification objectives can be very large and expensive.

Stork & Gill [26] have designed a specific diffractive optics to maximize the performances of what they called the “PicoCam”: a device made of a DOE and sensor array of less than 500µm in thickness, roughly 1mmx1mm, lightweight (30µg) and costing only a few cents.

In that case, the “star shape” of the grating used (Figure 19) is not the result of a global numerical optimization but comes from geometric considerations to reach enhanced signal to noise ratio on very thin systems.

![Figure 19: PicoCam principle (from [26]). A specific “star shape” gratings (a), leading to PSF with “rotating shape” (b). The imaged raw data (d) is processed with computational imaging (e) to recover the input data (c).](image1)

The above design gave birth to the company “ScoutCam” ([https://www.scoutcam.com/](https://www.scoutcam.com/)) using these very small cameras for all sorts of complex endoscopy situations.

In microscopy also, the concept of “lensless” imaging enables to design high-resolution, high field of view, low-cost and portable systems [27].

The very general principle is to use the sample under observation as a diffractive object (usually in transmission) and to image the interference pattern with a sensor array a few hundreds of microns below the sensor (Figure 20).

![Figure 20: Concept of lensless microscopy (from [27]). i) View of lensless microscope with LED light source (A), sample plan (B) and detector array (C). ii) The result of a large field of view imaged with the microscope, when a classic 40x field of view can be less than 500µm. iii) View of portable lensless microscope for air quality analysis.](image2)

Once imaged, a phase unwrapping algorithm on the interference pattern is used to reconstruct the intensity and phase image of the wave front. It is even possible to couple the unwrapping algorithms with super-resolution and diffraction gratings as mentioned in the previous section to enhance the system performances.

There are numerous applications to follow large size biological samples with a large magnification. Another use-case of lensless microscopy described in [28] is a low cost, portable device (~600g, <$150) to analyze the quantity and size distribution of air polluting particles (Figure 20).
c. 3D imaging

There are several ways to come-up with 3D imaging. The first one is to use multiple cameras with slightly shifted view angles, this is classically called “stereo-vision”. Another possibility is to capture the several images of the scene with variable focusing plane and a small depth of focus. The ability to characterize in which image each pixel is the “ sharpest” provide an idea of its distance to the focal plan. Time of flight cameras and laser scanner or lidar use a measurement of the light propagation time between the camera and the target. Plenoptics cameras or light-field imager use sets of micro-lenses behind the main camera optics to reconstruct virtual focal plane and get 3D information and enhanced depth of field using computational algorithms, at the cost of decreased resolution and lower signal to noise ratio [29], [30]. Raytrix is a company producing commercial plenoptics systems (https://raytrix.de/).

Another possibility to infer 3D information from a single image is to use the “depth from defocus” approach. The more an object is far from the camera object focal plan the more blurred it appears in the resulting image (Figure 21), it is therefore possible to analyze the amount of “blur” of an object to estimate its depth. It is even possible to computationally “refocus” the image to correct the blur effect (Figure 22).

![Figure 21: Point Spread Function as a function of the focal plan distance (from [3]). The point spread function can be considered as increasing homothetically with the distance to the focal plan.](image1)

![Figure 22: Example of 3D and extended depth of focus imaging using aperture coding and computational imaging (from [31]). Zoom on input raw data compared to processed “all-focus” images illustrates the deblurring effect.](image2)

If the PSF is known, it is possible to “invert” the blurring effect by deconvolution. To infer 3D position of an object, several PSF corresponding to different depth can be tried, and the dimension of the PSF providing the sharpest image after deconvolution gives an information about the depth. This is a classical
procedure, however, with regular refractive optics and continuous PSF, the result is not very accurate. Computational imaging enable to optimize the PSF function (and its dependency to the distance to the focal plan) for 3D reconstruction using diffractive optics [32][33]. See also an interesting qualitative explanation in [34] why specific PSF (like the one presented in Figure 19) are better than others for 3D reconstruction.

d. Phase imaging

Phase imaging is another field that can benefit from the latest technology in computational imaging. Phase imaging is used in many areas like laser metrology, high-precision 3D measurement [35], microscopy (see previous section) or adaptive optics (compensating the effect of atmosphere optical heterogeneities for instance). There are several technologies based either on Shack-Hartmann principles or interferometry [36].

In a recent paper, Wu et al. proposed a new approach based on computational imaging with SLM [37]. The key idea is to use several successive random configurations in the SLM to acquire several images with modified phase shifts. A dedicated (Gerchberg-Saxton) algorithm is then used for phase retrieval leading to strong enhancement in reachable resolution compared to classic Shack-Hartmann devices.

![Figure 23: Principle of WISH phase imager (from [37]). Top: experimental setup, successive acquisition with various SLM masks, intensity and phase resulting images of a fingerprint. Bottom left: acquisition schematics. Bottom right: potential use for imaging through scattering media: the transmission matrix from the scattering media can be measured through the phase measurement obtained with WISH and inverted using computational techniques.](image)

These applications are enabled by modern SLM that can be used with high frequency rate leading to fast phase mask modifications and overall acquisition speed for the device of about 10 Hz (each acquisition being itself composed of 10 shots with different phase masks). Computational imaging is used to define the best “pseudo-random” phase shift patterns produced by the SLM, combined with the optimization of the phase reconstruction algorithms that can then be used for all applications mentioned above, including adaptive optics and correction of images acquired through a scattering media (Figure 23).

e. Spectral imaging

Spectral imaging consists in enlarging color imaging by getting a full light spectrum (light intensity as a function of wavelength) for each pixel of an image instead of only three channels (usually red, green and blue). The measurement can be done in the visible range (wavelength from roughly 400 to 800 nm) or in any kind of spectral range depending on the availability of suitable detectors. Infrared spectroscopy as a lot of interesting applications in agriculture, biology, material sciences, remote sensing and many others since the measured spectrum in MWIR (~3-5µm) and LWIR (~8-12µm) reveals a lot of information about the molecular structure of imaged chemical species and/or their temperature ranges.
Spectral imaging enables those to obtain a spatial mapping of various chemical compounds. Several companies like Specim or Telops propose spectral imager based either on the “pushbrum” principle where dispersive optics (a prism) spreads the light as a function of its wavelength (like water droplets in a rainbow) or interferometric principle where a moving mirror is used to create an interference pattern that depends on the wavelength of the incident light. Both technologies involve some sort of scanning to obtain a real 2D spectral image (spatial scanning in the first case, temporal scanning of the moving mirror in the second case), which can be very detrimental for some fast applications. There is also an emergence of “snapshot” hyperspectral imaging, specifically in the visible and near-infrared wavelength range (NIR: 800nm–1µm) where sensor array can still be made of chip silicon semi-conductors [38].

The evolution of diffractive optics and computational imaging principles made it possible to design spectral imager with enhanced capabilities and low-cost optics. In [39], an end-to-end (joined optimization of optics and reconstruction algorithm) framework is presented for developing a 3D hyperspectral imager made on single snapshot imaging (Figure 24).

If the concept presented in this paper is close to several examples presented in the previous sections, there is still an interesting specificity which is the lack of availability of a 3D and hyperspectral “training set” to tune the system and train the deep-learning network used as the computation layer to retrieve spectral and 3D information from a single snapshot imaged through diffractive optics (Figure 25). An important work presented in the paper was to build reference data that could be used for experiments. This has been done using reference spectrometer and structured light 3D acquisition reference for 18 scenes and 73 reference objects. This application demonstrates the strong interest of computational imaging in terms of device cost. A regular reflex is transformed into a 3D hyperspectral imager with only the cost of an additional DoE, which is much less expansive for both prototyping and industrial production than high-quality dispersive refractive optics like prisms. This is clearly a breakthrough for hyperspectral applications in many areas where rentability costs are a key driver.

If spectral imaging in the visible range has many immediate applications, there are also a strong interest of spectral imaging for industrial or defense applications with longer wavelengths in the infrared. Unfortunately, detectors, but also refractive optics (lenses), become much more expensive (when they are even manufacturable). Diffractive optics and meta-lenses become also a solution for making spectral imaging at larger wavelengths much more affordable. This is a strong focus (among other computational imaging problem) of Menon’s team in the University of Utah which has been working actively to design
multi-level diffractive flat lenses suitable for Long Wave Infra-Red imaging [40] (LWIR: ∼8-12 µm in wavelength).

In the LWIR wavelength range, Sullenberg et all [41] have recently proposed a strategy like compressed sensing or aperture coding using a ZnSe binary encoding mask (see section 1.c) to design a spectral imager for this range of wavelength without active components. The first successful flight tests with this instrument have been conducted and results published in [42].

Figure 26. From [42]. a) Principle of CRISP spectral imager. The scene (b) is imaged through dispersing prisms and a ZnSe binary aperture encoding mask (c). The spectral information related to the scene (d) can be retrieved from multiple images acquired by a plane or drone (e) using computation like those described in section 1.c.

THz imaging

TeraHertz imaging has made a lot of progresses in the last years, thanks in particular to the development of more powerful sources and more sensitive detectors [43]. TeraHertz are electromagnetic waves with wavelengths ranging from 0.1 to 3mm (between the infrared and microwaves). It is widely used in airports for security purposes since it enables to “see” metal through clothes (both metal and the water in human body are opaque while fabric is transparent), but it has many more applications since it can penetrate quite far in many materials.

Traditional optics are very hard to manufacture for THz, so the aperture coding framework and computational imaging has a wide range of applications in this range of wavelength. One of solutions to get images with TeraHertz is based on the “single pixel” camera paradigm, closed to compressed sensing. The principle is “aperture coding” in which binary masks enable to superpose shifted version of the same scene on a sensor array (see section 1.c). However, to solve the inverse problem in aperture coding, several measurements are required. In classical situations, the multiple pixels of the detector array are used as multiple measurements, whereas in the “single-pixel camera” framework, successive acquisitions with the same single detector are used as multiple measurements. Of course, these measurements must be as independent as possible to maximize the information we get from them. In THz imaging, the source of variability in the temporal domain usually comes from the phase modulation of the THz source. In [44] and [45], deep-learning is used in a computational imaging fashion to speed-up and improve solving the aperture coding inverse problem based on THz modulated sources.
In [46], a different aperture coding framework is described, closer to traditional single-pixel optical cameras: the aperture coding mask is modulated thanks to a silicon prism that can be photo-activated with a matrix of micro-mirrors (Figure 27).

![Figure 27](image)

Figure 27: Aperture coding system for THz single-pixel camera (from [46]). The system is used in transmission configuration with synchronized THz light source, THz detector and a matrix of micro-mirrors acting as time-varying aperture coding mask.

Thanks to the speed of modulation with the micro-mirror array (~20kHz), a THz video with 32x32 pixel resolution and 6 frames/second can be achieved with a single-pixel detector.

g. Conclusion and synthesis

Lots of practical applications for the global paradigm of computational imaging can be found in the recent literature from improving form factors of camera systems to overtaking state-of-the-art performances in the field of wave-front sensing or 3D spectral imaging.

All these applications are enabled on the one hand by the evolution of optical technologies (some of them presented in the previous section), the decrease in cost and form factors of various processors and computing architecture such as CPUs, GPUs or FPGAs and the global concept of computational imaging in which collection optics and algorithms are considered as two inter-dependent parts of a single “optical system”.

Below are a few key principles underlying computational imaging that can be drawn from the previous examples:

- New optical compounds are very flexible and can be designed with many degrees of freedom. But fixing those parameters in the most efficient way can involve complex optimization procedures.
- Aperture coding is a very generic framework that can be adapted to many situations.
- Cost functions to be optimized or problem constraints should be considered with care. Real financial cost, form factor, robustness, optical behavior related to intensity, phase or polarization can be suitable candidate terms in the global cost function to optimize in an end-to-end fashion, but also in a more classical and sequential manner. It is important to consider those terms in relation with the real practical use of the final optical device.
- If possible, prior knowledge about the scene or object to be imaged should be used. This is sometimes necessary when complex inverse problems must be solved using regularization terms or when machine learning is used based on pre-defined training sets. In all those situations, care should be taken over the relevancy of generalizing the considered priors or over the representativity of the training set.
CONCLUSION - LIMITATIONS AND PERSPECTIVES

CONCLUSION
The target of this state-of-the art report was to show how the latest developments in optical design, nanotechnologies, simulation and machine learning algorithms could unlock new areas of developments and new applications for imaging systems.

Many economic reports forecast a strong growth in the “computational imaging” market in the coming years (MarketsandMarkets: 22% of growth from $10.7 billion to $29 billion in 2024, Persistence Market Research: annual growth of 7.6% for DoE market in the 5 coming years…), mostly because the demand for high-performance, low-cost, miniaturized, energy-efficient, autonomous embedded optical components is strongly growing in applications such as smartphones, telecommunications, sensing in the Internet of Things context, drones, biological and medical imaging but also virtual reality, LIDAR for autonomous vehicles and many others…

An important factor to consider is the ratio between imaging device costs and economical added value. A very important driver for the evolution of imaging systems is the strong decrease of manufacturing costs for many modern optical components such as CMOS sensors, DoE and metalenses thanks to the progresses made in material sciences and nanotechnologies in general. Being able to replace expensive refractive optics on the one hand by chipper components completed by computing power and smart algorithms on the other hand could be a real breakthrough.

Nevertheless, many of the technologies presented above are presently at the laboratory development stage. For instance, the start-up Tunoptix has just been awarded in 2020 a 223k$ grant by DARPA to develop metasurface lenses… The market growth today remains therefore at the perspective level and there are still a lot of progresses to be made in industrializing and producing metalenses, diffractive optics or complete computational imaging systems at large scale.

LIMITATIONS
From the technical point of view, there is always a risk when training a system for a dedicated target by integrating knowledge about the expected result to improve the output of the system. When using the system in a day-to-day fashion, some safeguards should be implemented to make sure that there is no drift in time between the acquired scenes and the priors (or training set) used to design and optimize the system. In some way, using tools such as deep-learning or regularization techniques with strong assumptions related to the object to be imaged should be completed by a “validity domain” for the expected inputs. But such a domain can be difficult to define leading to potential difficulties for assessing the effective precision of the device over the time.

Another possible concern is that end-to-end design is somehow limiting the compatibility between optical components and their ability to be re-used in other applications. In current optical system, it is often possible to use optics with multiple cameras or for multiple purposes. If both DoE and algorithms are optimized together for a dedicated situation, maybe critical mass markets will be more difficult to reach because of over specialized devices.

PERSPECTIVES
Optics is a very dynamic field of research in many areas that were not mentioned here like laser and light sources, up-conversion devices, semi-conductor improvements and many others. There are therefore many possibilities to use the very fruitful computational imaging paradigm with much more complex optical compounds than those presented here with many perspectives in improving intrinsic optical sensing capabilities.
APPENDIX 1 – Glossary

**Aperture coding**: One of the most classical “computational imaging” example. A mask with random openings is placed in front of the collection optics leading to interesting image capabilities if the mask is well chosen together with the deconvolution algorithms.

**BRDF**: Bidirectional Reflectance Distribution Function. It is a classical way to encode the optical behavior of a surface at macroscopic scale (specular and diffuse reflectivity). It is used a lot in ray tracing or geometric optics simulation. A perfect black body will have a constant brdf (light diffused equally in all directions), while a perfect mirror will be described by a Dirac brdf.

**CCD**: Charged Coupled Device is a common technology to build light sensitive matrix. CCD and CMOS differ mostly by the amplification and reading principle of the number of charges collected by the semi-conductor pixels.

**CMOS**: Complementary metal-oxide semi-conductor is a technology that can be used to produce light-sensitive sensor arrays.

**CNN**: Convolutional neural networks: A specific neural network within the deep-learning framework for working with images as input. Networks are built so that the linear part of the neural-network can be expressed as convolution functions. The training process mostly tunes the convolution kernel iterations after iterations.

**CPU**: Computer processing unit. This is the classical type of processor embedded in computers.

**Deep-learning**: A specific part of machine learning where models with a lot of fitting parameters are used to solve complex problems. Deep-learning algorithms usually have the capability to approximate complex functions but requires a lot of training data.

**DoE**: Diffractive optical elements. Usually thin transparent media in which the thickness can vary in a controlled manner from one point to another leading to specific phase shifts and diffraction patterns.

**End-to-end optimization**: Procedure that consists in trying to find a global optimum for both optical configuration and image processing. In the document, the terms “joined optimization” has been used with the same sense.

**Fresnel lens**: Specific type of lens in which a thick refractive optics is replaced by a thin discontinuous surface, leading to the same phase shift and convergence properties. It is usually difficult to make high quality Fresnel lenses for a large wavelength range.

**GPU**: Graphical processing unit. Specific type of processor usually embedded on graphics card for computer vision which have high capabilities for parallel computing and which are therefore used a lot for machine-learning.

**MEMS**: Micro-electro-mechanical systems are usually small components based on silicon where electronics are intimately coupled with mechanical functions (moving parts for instance).

**Meta-lenses**: Optical components like DOE, but with nanostructure imprinted at specific location in order to give more control on the light beam (phase shift and polarization).

**MOEMS**: Micro-opto-electro-mechanical systems. Like MEMS, but with optical components interacting with photons included.

**Optical system**: In this document, optical systems are defined as the set of all optical components AND image processing algorithms leading to the creation of a digital image.

**SLM**: Spatial light modulator. These devices usually made of liquid crystals enables to modify in real time the spatial phase shift of the light reflecting or crossing the device (this is in some way like DoE, with the possibility to tune and change the modulation function on demand).

**SPAD**: Single Photon Avalanche Diodes. Array of photo-detectors with the particularity to be sensitive to very low light level with ultra-small exposure time and fast acquisition rate.
FIGURE 28: Spherical wave diffracted by a DOE.

Figure 28 represents a classical geometric configuration of a diffractive optics located at a distance $z$ of a point source $P$ (all developments are in 2D, but generalization to 3D is straightforward). The diffractive optics has a thickness $e$, and a sensor array is located at a distance $f$ of the optics. The Point Spread Function $P(P, v)$ represents physically how the light intensity from point $P$ is spread on various pixels of the sensor array in the plan $Z = z + f$. The expression of $P(P, v)$ associated with a diffractive optics made of material with refractive index $\eta$ at wavelength $\lambda$ can be approximated by relatively simple expressions [8].

If the electric field $U_\lambda(x, y)$ corresponds to a spherical wave coming from $P$, then the electric field in the plan of the DOE can be expressed as:

$$U_\lambda(x, z) = A e^{-\frac{i2\pi}{\lambda} r(M)} \approx A e^{-\frac{i\pi x^2}{\lambda}}$$

This is only a valid approximation if $z$ is much larger than $x$, which is usually the case. This expresses simply the phase shift at the different location of the DOE, which depends on the distance between $P$ and $M$. $A$ is a constant factor that does not depend on the position $x$.

Then if the DOE is thin enough, meaning that its thickness is of the order of magnitude of the wavelength, it can be considered that the DOE effect is simply a phase shift that depends on its local thickness $h(x)$. However, if some topographical structures have spatial wavelengths smaller than the wavelength of the light and significant amplitudes, then other diffusion phenomena must be considered (this would be the case for metalenses for instance), and if the DOE is too thick, then the geometrical approximation of an “infinitely thin” surface is also not valid.

But if $e \sim \lambda$, then the expression of the electric field at the exit of the DOE can be expressed simply:

$$U_\lambda^3(x) = U_\lambda(x, z) e^{-\frac{i2\pi}{\lambda} (\eta\lambda - 1) h(x)} \approx A e^{-\frac{i\pi x^2}{\lambda} + 2(\eta\lambda - 1) h(x)}$$

Then, the computation of the electric field at point $Q$ can be obtained with a Rayleigh-Sommerfeld integration over the surface of the DOE:

$$U_\lambda^2(v) \propto \int_{DOE} U_\lambda^3(x) \left( i \frac{2\pi}{\lambda} - \frac{1}{r(x, v)} \right) e^{\frac{i2\pi r(x,v)}{r(x,v)^2}} dx$$
\[ U^2_\lambda(v) \propto A \int_{DOE} \left( i \frac{2\pi}{\lambda} - \frac{1}{r(x,v)} \right) \frac{1}{r(x,v)^2} e^{-\frac{\langle x \rangle^2}{\lambda^2}} e^{-i\pi(\eta\lambda - 1)h(x) + 2r(x,v)} dx \]

This expression can be simplified further by approximating \( r(x, v) \) as a function of \( v \) and \( f \) if \( f \) is much larger than \( v \). The point spread function \( P(M, v) \) is simply the average value of the electric field at point \( Q(v) \). It can then be used to integrate over the entire space to compute the image of a complete scene through the diffractive optics.

Depending on the considered approximations, \( P(M, v) \) can be expressed with a simple and differentiable expression of the height field \( h(x) \) which represents the local thickness of the DOE, making it suitable for optimization algorithms.

ACKNOWLEDGEMENT AND READING NOTICE:
I would like to thank the Ecole Polytechnique Executive Master team for letting me the opportunity to take a deeper look at these use-cases of diffractive optics and computational imaging, and I would like also to thank Grace Ham, scientist at ArcelorMittal Maizieres Research, who first attracted my attention to those great topics, which, as surprising as it may sound, can probably find great applications for developing new measurement systems on steel production lines.

The above document intends mostly to provide a wide overview of the foundations and practical use of computational imaging with enough references on which the interested reader could rely to find more accurate scientific details. Equations and mathematical formalism are maybe not as rigorist as they should be, but I found them nevertheless helpful to express and understand the underlying mathematical and physical principles sustaining the computational imaging framework, at least in a qualitative fashion.


[38] ‘Commercial Snapshot Spectral Imaging: The Art of the Possible’, p. 42.


