HAL
open science

Comparison of methods for end-to-end co-optimization of optical systems and image processing with commercial lens design software

Alice Fontbonne, Hervé Sauer, François Goudail

To cite this version:

Alice Fontbonne, Hervé Sauer, François Goudail. Comparison of methods for end-to-end co-optimization of optical systems and image processing with commercial lens design software. Optics Express, 2022, 30 (8), pp.13556. 10.1364/OE.455669 . hal-03636740v2

HAL Id: hal-03636740

https://hal.science/hal-03636740v2

Submitted on 19 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of methods for end-to-end co-optimization of optical systems and image processing with commercial lens design software

Alice Fontbonne,[1],* Hervé Sauer,[1] and François Goudail [1]

[1] Université Paris-Saclay, Institut d'Optique Graduate School, CNRS, Laboratoire Charles Fabry, 91127, Palaiseau, France
*alice.fontbonne@protonmail.com

Abstract

We compare three different methods to co-optimize hybrid optical / digital imaging systems with a commercial lens design software: conventional optimization based on spot diagram minimization, optimization of a surrogate criterion based on a priori equalization of modulation transfer functions (MTFs), and minimization of the mean square error (MSE) between the ideal sharp image and the image restored by a unique deconvolution filter. To implement the latter method, we integrate – for the first time to our knowledge – MSE optimization to the software Synopsys CodeV. Taking as application example the design of a Cooke triplet having good image quality everywhere in the field of view (FoV), we show that it is possible, by leveraging deconvolution during the optimization process, to adapt the spatial distribution of imaging performance to a prescribed goal. We also demonstrate the superiority of MSE co-optimization over the other methods, both in terms of quantitative and visual image quality.

This accepted author version has been published as: Alice Fontbonne, Hervé Sauer, and François Goudail, "Comparison of methods for end-to-end co-optimization of optical systems and image processing with commercial lens design software," Opt. Express 30, 13556-13571 (2022)
DOI : https://doi.org/10.1364/OE.455669
Supplemental document 1 is provided at the end of this pre-print as Annexe A.

1 Introduction

Currently, almost all imaging systems integrate both a complex optical system and digital processing algorithms. The idea of modeling the entire imaging chain to enable its optimization is not new, but it took several years before the first co-design approaches really appeared. The first to co-design an optical system were Cathey & Dowsky [10, 2], focusing on the co-design of a single optical element, a phase mask. The next step is to take into account more

complex optical systems and to co-optimize them with digital post-processing algorithms. These methods are referred to as "joint optical-digital design"[1], "integral design strategy"[36], "end-to-end lens design"[31] or "holistic optical-digital hybrid-imaging design"[35] in the literature.
Recent approaches attempt to bypass the expertise of commercial optical designers by using deep learning techniques for co-design [28, 25, 19, 31, 17, 14]. However, currently, only commercial optical lens design software (like Zemax® OpticStudio® or Synopsys® CodeV®) allows the design of complex optical systems that are truly manufacturable. This is why it is interesting to integrate directly a co-design approach into this type of software. Indeed, the optical designer expertise allow them to use many hints in order to steer the lens optimization process [12], which usually consists of a sequence of local optimizations towards evolving intermediate goals to reach a final realistic design with good final image quality.
This idea of integrating a co-design metric taking into account the entire hybrid system, namely, the mean square error (MSE), is not new in Zemax. Stork & Robinson [30] were the first to implement it, followed by Vettenburg & Harvey [35] and Wang et al. [36]. However, to the best of our knowledge and up to now, no publication claims to have implemented it in CodeV. Moreover, the results obtained with a MSE-based criterion and a commercial lens design software have never been compared to other co-design methods. Indeed, there are several methods to optimize a hybrid system from end-to-end. In particular, another joint optical-digital design method has been developed by Burcklen et al. [1] to facilitate the use of the co-design paradigm by professional optical designers in CodeV. It is based on a "surrogate" merit function that contains terms classically available in lens design software but used in a non-standard way to implicitly take into account image processing.
Our purpose in this work is to implement the MSE-based optimization criterion in the CodeV software, and to compare the obtained performance with two other optimization criteria. The example of application chosen for illustration is the design of a Cooke triplet aimed at providing good image quality over a wide field of view (FoV) of 40° (±20°). The basic elements of the considered hybrid optical-digital system are described in Sec. 2. In Sec. 3, we apply the "conventional" design method, which considers the existence of the digital processing only after the optimization of the optical system. We establish the performance of the obtained hybrid system by evaluating its modulation transfer functions (MTFs), its effective MTF (taking into account the whole optical system and post-processing) and the visual quality of the restored images. In Sec. 4, we investigate in the same way the method based on a surrogate criterion implemented in Code V [1]. Then, in Sec. 5, we describe the implementation of a method for directly optimizing the MSE under CodeV, and compare the performance of the obtained hybrid system with those provided by the two other methods. We conclude in Sec. 6 on the performance of the three studied methods and on the most efficient way to take into account the existence of a digital post-processing during the lens design process.

2 Description of the hybrid system of interest

This section aims to describe the basic specifications of the optical/digital hybrid systems considered in this article. Their goal is to obtain the best possible image quality over the whole FoV. A hybrid system is made of two blocks: the imaging system (i.e. the optical system and a sensor) and a digital post-processing algorithm. For the purpose of this study, we have chosen to consider a classical lens architecture, a Cooke triplet (described in Sec. 2.1), and a fast deconvolution algorithm, the average Wiener filter (described in Sec. 2.2). The optimization goal and the different investigated methods to achieve it are described in Sec. 2.3.

2.1 Optical architecture

The original purpose of a Cooke Triplet was to serve as a photographic lens. Consisting of three optical elements and a diaphragm, it is recognized for its good correction of aberrations over a wide FoV [32]. Its variants have been the subject of several studies [26, 27, 34] and it is still used today for the improvement of optical systems [29]. The Cooke triplet is thus a basic "building block" of conventional optical design, which is interesting to study in the context of joint optical/digital design. We will thus use it in this article as a basis for the description of the various optimization methods. We choose to set its focal length to 50 mm, its aperture to F/4 and its half-FoV to 20° (it is enlarged compared to the initial 14° half-FoV of the CodeV example from which this triplet is derived).
We will consider that this lens is used with a panchromatic sensor of pixel size about 5 µm, which sets the Nyquist frequency to $f_N = 100$ lp.mm^{-1}. Three wavelengths of the visible spectrum are considered, and the constraints applied to the system during optimization are standard (see Supplemental 1, section 1).

2.2 Post-processing algorithm

The digital post-processing applied to the image acquired by this system consists of a deconvolution to compensate for the aberrations of the optical lens. We want the deconvolution algorithm to ensure reconstruction regardless of the position in the FoV and to be fast and simple. For this purpose, we choose the linear average Wiener filter:

$$\tilde{w}_\Psi(\nu) = \frac{\frac{1}{K}\sum_{k=1}^{K} \tilde{h}_{\psi_k}(\nu)^\star}{\frac{1}{K}\sum_{k=1}^{K} |\tilde{h}_{\psi_k}(\nu)|^2 + \frac{S_{nn}(\nu)}{S_{oo}(\nu)}} \quad , \tag{1}$$

where $\Psi = \{\psi_1, \psi_2, ..., \psi_K\}$ is a set of K positions in the FoV describing the whole FoV, $\tilde{h}_{\psi_k}(\nu)$ is the local optical transfer function (OTF) of the imaging system at the position ψ_k depending on the spatial frequency ν, $S_{nn}(\nu)$ is the power spectral density (PSD) of the noise and $S_{oo}(\nu)$ is the statistical PSD of the scene. We use a generic ideal image model with a power-law PSD [23, 33] $S_{oo}(\nu) \propto \nu^{-2.5}$, which well represents natural scenes. The white noise

PSD is such that the signal-to-noise ratio (SNR) on the raw image is 34 dB, with $SNR = 10\log_{10}\left[\int S_{oo}(\nu)d\nu / \int S_{nn}(\nu)d\nu\right]$. Note that in order to obtain rotationally symmetrical average Wiener filters, it is better to use a large number of positions in the FoV. In the following, unless otherwise stated, $K = 400$. Furthermore, in view of the expression of the Wiener filter in Eq. 1, we can define the "average MTF" as:

$$\sqrt{\frac{1}{K}\sum_{k=1}^{K}|\tilde{h}_{\psi_k}(\nu)|^2} \quad . \tag{2}$$

2.3 Optimization goal

Our objective is to design a hybrid system with good image quality over the whole FoV. To reach this goal, we will investigate three different design approaches.

- The "conventional triplet" is obtained with a conventional lens design criterion based on the minimization of the average size of the spot diagram over the FoV. This method does not take into account the digital post-processing during the optimization process (Sec. 3).

- The "EMTF triplet" is obtained by optimizing the surrogate criterion based on conventional optical metrics introduced in [1]. It implicitly takes into account the post-processing (Sec. 4).

- The "MMSE" triplets are obtained by minimizing the average MSE between the deconvolved image and an ideally sharp image of the scene. More precisely, the image produced on the sensor at one position ψ_k of the FoV can be modeled by $h_{\psi_k}(r) * O(r)$, where $O(r)$ is the sampled ideal scene image (r represents the spatial coordinates) and $*$ denotes the convolution operator. This acquired image is then deconvolved with the filter $w_\Psi(r)$, defined in Eq. 1 in the Fourier domain, to restore its sharpness. The restored image at the output of the hybrid optical-digital system can thus be modeled as:

$$\hat{O}_{\psi_k}(r) = w_\Psi(r) * [h_{\psi_k}(r) * O(r) + n(r)] \tag{3}$$

  where $n(r)$ is the detection noise. The mean-squared error (MSE) is then defined as:

$$MSE(\psi_k) = \mathrm{E}\left[\int \left|\hat{O}_{\psi_k}(r) - O(r)\right|^2 dr\right] \quad , \tag{4}$$

  where $E[\cdot]$ represents the mathematical expectation over the noise $n(r)$ and the scene image $O(r)$, which are both assumed to be zero-mean, stationary random processes of PSD $S_{nn}(\nu)$ and $S_{oo}(\nu)$ respectively. The MSE Eq. 4 can in fact be generically expressed in the Fourier domain by

the sole noise and statistical scene image PSDs, the OTF $\tilde{h}_{\psi_k}(\nu)$ of the lens and the restoration filter $\tilde{w}_\Psi(\nu)$ [21, 8]:

$$MSE(\psi_k) = \int \left(|\tilde{h}_{\psi_k}(\nu)\tilde{w}_\Psi(\nu) - 1|^2 S_{oo}(\nu) + |\tilde{w}_\Psi(\nu)|^2 S_{nn}(\nu) \right) d\nu \quad . \quad (5)$$

The lens optimization criterion is then defined as the MSE averaged over all the considered FoV positions $\Psi = \{\psi_1, \psi_2, \cdots, \psi_K\}$:

$$MSE_{mean}(\Psi) = \frac{1}{K} \sum_{k=1}^{K} MSE(\psi_k) \quad (6)$$

Note that the average Wiener filter defined in Eq. 1 is the linear deconvolution filter that minimizes MSE_{mean} for a given optical system [8], which leads to overall optimizations in the MSE_{mean} sense. As all (local) optimization methods, its result depends on the starting point. We will consider two different systems: the "MMSE_1 triplet" will be obtained by taking the "conventional triplet" as starting point, and the "MMSE_2 triplet" will take the EMTF triplet as a starting point (Sec. 5).

In the remainder of this paper, we will describe the optimization of hybrid imaging systems with these three different approaches and compare the obtained performance.

3 Conventional approach

3.1 Definition of the conventional approach

To limit the aberrations in an optical system for a selected number of positions in the FoV, the conventional optimization criterion in CodeV is to minimize the sum of the squares of the root mean square (RMS) diameters of the spot diagrams (calculated from a set of desired positions in the FoV). Other criteria exist natively in CodeV: it is possible to minimize the wavefront error (WFE) or to maximize the value of the modulation transfer function (MTF) at selected spatial frequencies. Whatever the chosen criterion, it is always necessary to impose constraints on some parameters during the optimization (for example, set the focal length to a particular value defined by the specifications, or constrain the thickness at the edges of the lenses to have strictly positive values). This can be done by adding penalization terms like $p^2 \times (f' - f'_{target})^2$ (where p is a positive constant, defining the "weight" of this term) to the basic criterion, or by strictly enforcing the constraint using Lagrange multiplier. In this study, we optimize the sum of the squares of the RMS diameters of the spot diagrams under common CodeV optical constraints (see Supplemental 1, section 1).
The CodeV optical design software uses by default the Levenberg-Marquardt (or Damped least squares) algorithm to minimize this cost function [16, 18, 9, 20], but other algorithms can be used or implemented [22, 15, 24]. Since the

Levenberg-Marquardt algorithm is a local minimization algorithm, convergence to a global minimum is not guaranteed. The expertise of a lens designer is therefore necessary, both to choose a suitable starting point for the optimization [6] and to steer it towards an acceptable local minimum. Here, the starting point is part of the CodeV example library. The triplet is optimized in the CodeV conventional way by setting as variable all element surface curvatures, element center thicknesses, element distances and also indices (n) and Abbe numbers (V) of glass materials, letting general constrains (see Supplemental 1 Section 1) warrant that the system is physically realistic. Once the optimization is done, true existing materials close to the continuously optimized (n, V) values are chosen from the CodeV glass materials library, among the inexpensive and common glasses from the Schott and Ohara catalogs, and the system is re-optimized, resulting in the lens shown in Fig. 1(a) and precisely described in Supplemental 1 (section 2), referred to as the "conventional triplet".

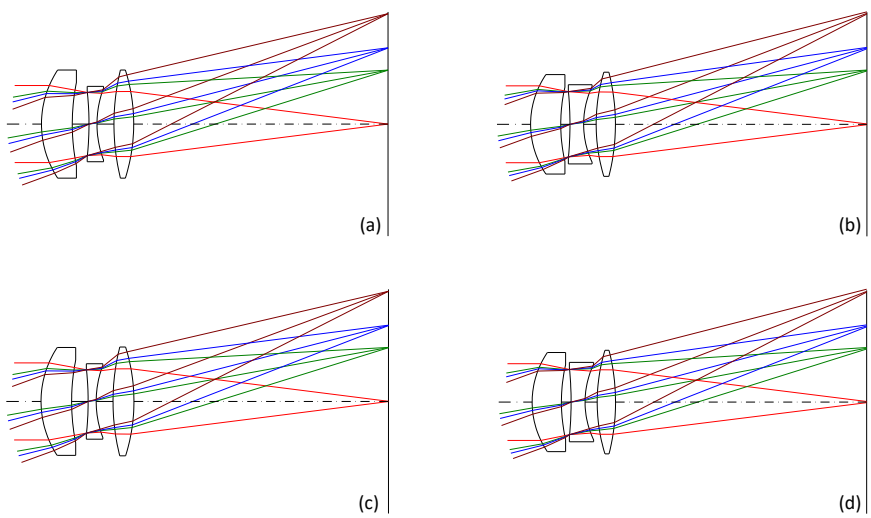

Figure 1: Scheme of Cooke triplets. (a) "Conventional triplet". (b) "EMTF triplet". (c) "MMSE_1 triplet". (d) "MMSE_2 triplet".

In order to evaluate the performance of a hybrid system based on this triplet, we will consider several evaluation metrics. In Sec. 3.2, we analyze the MTFs at different positions in the FoV. It is a purely optical metric, useful to compare the raw image qualities on the sensor plane given by the different Cooke triplets. In order to take into account the digital post-processing, we also analyze the effective MTFs (a metric that takes into account the optical system and the post-processing, which will be defined in the next paragraph). Finally, in Section 3.3, we evaluate the global system performance through image simulation.

3.2 Analysis of the MTFs and effective MTFs

Figure 2(a) represents the tangential and sagittal MTFs of the conventional triplet at different positions in the FoV. It has the standard characteristics of a lens of this type: its structure allows a good correction of aberrations, which makes possible to obtain, on the axis, a MTF taking fairly high values (higher than 0.3 at the Nyquist frequency). However, performance degrades as the FoV increases: at 14° half-FoV, the sagittal MTF presents nullings at the spatial frequency 40 lp.mm^{-1}. This means that at (and around) this frequency, for this particular FoV position and orientation, the signal is drowned out in noise and cannot be recovered. The relative frequency-wise noise level (i.e. $\sqrt{S_{nn}(\nu)/S_{oo}(\nu)}$) corresponding to a SNR of 34 dB is shown with a black dotted line in Fig. 2(a) .

However, the MTFs alone are not representative of the performance of the whole hybrid optical/post-processing system, since they do not take into account the image restoration with post-processing. A more representative metric of the entire hybrid system is the *effective* MTF defined as:

$$|\tilde{h}^{eff}_{\psi_k}(\nu)| = |\tilde{h}_{\psi_k}(\nu) \times \tilde{w}_{\Psi}(\nu)| \quad . \tag{7}$$

In an ideal hybrid system, the effective MTF should be uniformly equal to 1, meaning that the signal components at all spatial frequencies are perfectly restored by the deconvolution filter. Figure 2(b) represents the effective MTFs of the conventional triplet. We can observe that for the highest frequencies (between 85 lp.mm^{-1} and 100 lp.mm^{-1}), the level of all the effective MTFs drop compared to the corresponding MTFs. This is due to the fact that the Wiener filter makes a compromise (the best in the sense of the MSE) between signal reconstruction and deleterious noise reinforcement. On the contrary, for spatial frequencies lower than 45 lp.mm^{-1}, the effective MTF on axis is enhanced and exceeds 1, which generates, as we will see later, visual over-contrast. This is due to the fact that the on-axis MTF is higher than the "average MTF" (Eq. 2) to which the average Wiener filter is adapted. In consequence, this filter "over-compensates" the on-axis MTF. Furthermore, the MTF nullings are still present on the effective MTFs, since information is totally lost at these spatial frequencies and cannot be recovered by the average Wiener filter.

3.3 Analysis of final image quality

Let us now verify the observations made on MTFs and effective MTFs through image simulations. Figure 3 shows the scene used for these simulations. It is a large image, of 5144 × 5144 pixels, chosen to obtain a FoV of 40° (±20°) on the diagonal of the sensor, with a pixel size close to 5 µm. Several subparts are selected on the raw image, to check the reconstruction of the details at three different positions in the FoV: on axis, intermediary position and extremity. Simulation of the observation of this scene through the Cooke triplet is performed with the image simulation tool of the CodeV software, with 20 × 20 variable PSFs in the FoV, in order to take into account all the defects of the

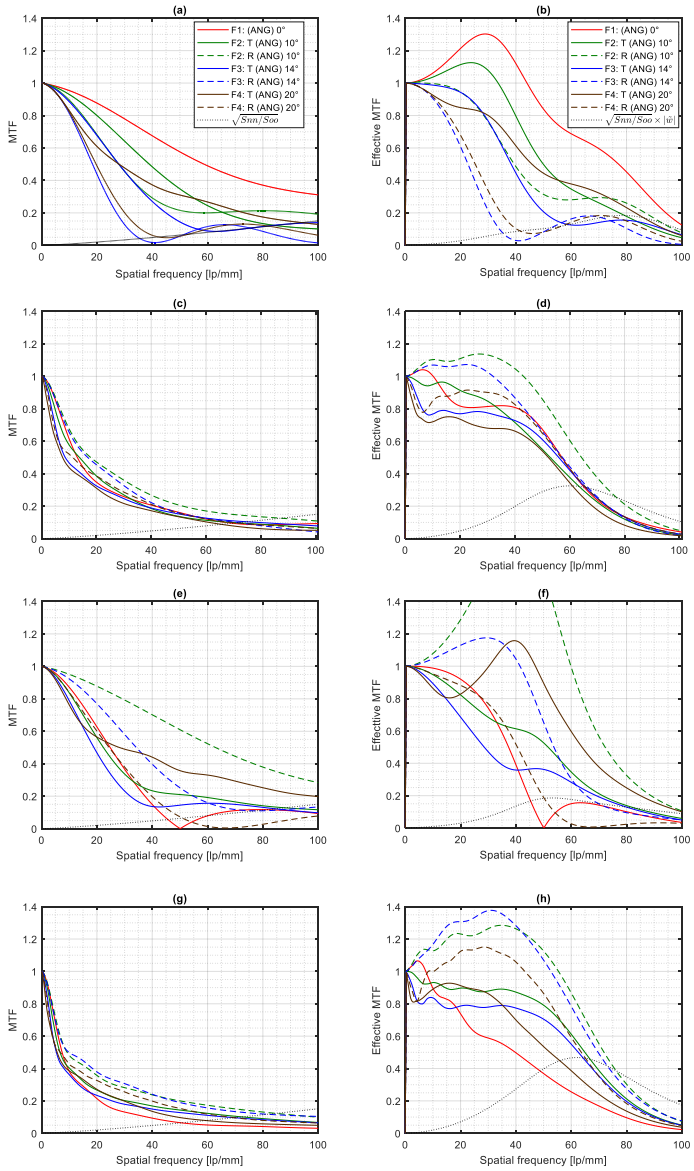

Figure 2: (First column) MTFs and (Second column) Effective MTFs of the (a-b) conventional triplet, (c-d) EMTF triplet (e-f) MMSE_1 triplet (g-h) MMSE_2 triplet. The curve legends for the all first column plots are the same as the one in (a) and the legends for the all second column plots are the same as the one in (b).

optical system. The image is then corrupted with a white Gaussian noise so that the SNR is 34 dB.

Figure 3: Image of the Gloucester Cathedral (public domain picture) used as a scene for the comparison of Cooke triplets optimized with different methods. Three images are selected for their details and their positions in the FoV (on axis, middle field, extreme field).

For the conventional triplet used without deconvolution, the resulting subpart images are shown in the first row of Fig. 4. One observes that the images appear blurred, especially those corresponding to large fields. Noise appears quite low, thanks to the high value of the SNR (34 dB). Figure 4(b) represents the same subparts after post-processing with an average Wiener filter constructed from OTFs corresponding to thirteen different positions in the FoV (see Supplemental 1 section 3 for more details). Deconvolution raises the contrast reduced by the optical system, and this can be shown theoretically with the mean image quality (IQ), expressed in dB and defined as

$$IQ_{mean} = 10 \log_{10} \frac{1}{MSE_{mean}(\Psi)} \quad . \tag{8}$$

It is a decreasing function of MSE_{mean} defined in Eq. 6. It thus increases

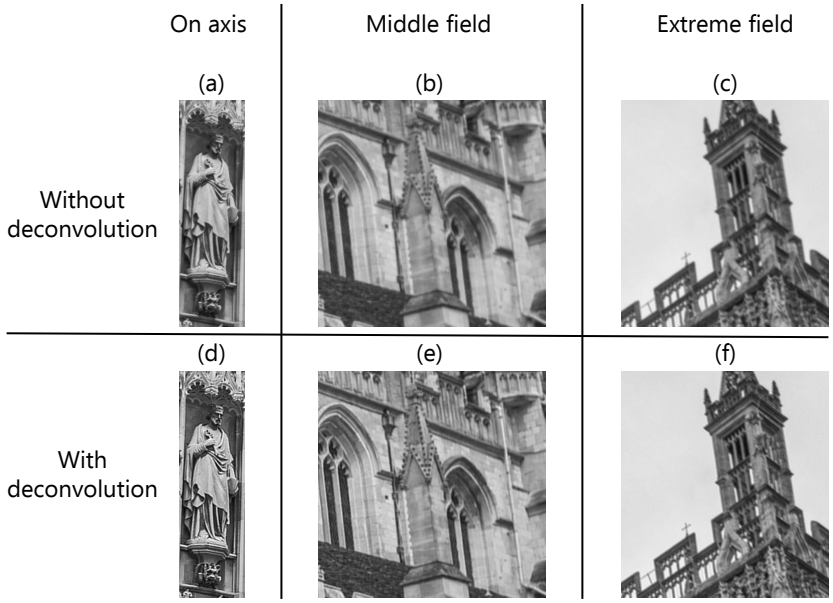

Figure 4: Details of the cathedral image obtained with the conventional triplet for an SNR of 34 dB, (a-b-c) without deconvolution and (d-e-f) with deconvolution.

IQ_{mean}	Conventional triplet	EMTF triplet	MMSE_1 triplet	MMSE_2 triplet
Before deconvolution	12.9 dB	10.9 dB	13.2 dB	10.1 dB
After deconvolution	13.5 dB	13.5 dB	14.6 dB	16.0 dB

Table 1: IQ_{mean} before and after deconvolution for the different Cooke triplets.

when the performance is enhanced. Table 1 contains the values of IQ_{mean} for all the Cooke triplets optimized in this article. For the conventional triplet, the IQ_{mean} goes from 12.9 dB up to 13.5 dB thanks to deconvolution. Nevertheless, this increase is not homogeneous across the FoV: the on-axis subpart image (Fig. 4(d)) is more contrasted than it should be (according to the ideal image in Fig. 3). The deconvolved middle field image (Fig. 4(e)) appears to be well reconstructed, while the deconvolved extreme field image (Fig. 4(f)) is only slightly more contrasted than the unprocessed subpart (Fig. 4(c)). These observations match the conclusions drawn from the effective MTFs. Indeed, Fig. 2(b), an effective MTF higher than 1 for low spatial frequencies indicates over contrast, while an effective MTF close to 0 indicates under contrast (blur).

4 Surrogate criterion based on optical metrics

4.1 Optimization method

Since our optimization goal is to obtain the best possible image quality over the whole FoV, it is possible to define the "best system" as having equivalent performance at any point of the FoV, after deconvolution. Since we have decided to use a single deconvolution filter, this is equivalent to optimizing the optical system under the two following general constraints.

- The MTFs at all the positions in the FoV should be nearly equal to each other, so that deconvolution by a unique filter yields equivalent performance across the FoV. To reach this goal, one should implement specific optimization constraints forcing the MTFs at all FoV positions to be as close as possible to each other for a few chosen spatial frequencies.
- The MTF values should be maximized to obtain the best possible frequency-wise SNR before deconvolution (and avoid any MTF nullings in the spatial frequency range of interest).

This approach will be called "co-optimization by MTFs equalization". It should be noted that the maximization of the MTF values is a very important point of this optimization technique. In addition to corresponding well to the goal of obtaining the best possible image quality over the whole FoV, this co-design criterion can be seen as an alternative (a "surrogate") to the MSE. Indeed, several studies have shown that optimizing performance in terms of MSE over a set of parameters (e.g. depths of field) leads to MTFs that are close to each other [11, 1, 13]. This criterion can be implemented in various ways in commercial optical software [1]. In this work, we have disabled the default CodeV error function based on the sum of squared spot diagram RMS diameters and only enforced the following user-defined constraints:

- Favoring equality between the MTFs at a set of some Y-meridional FoV positions and the on-axis MTF (it is a set of *weighted* constraints, to allow some latitude and balancing on these equalizations). In order to guarantee the similarity of the MTFs whatever the azimuth, one has to take into account the tangential, sagittal and 45° MTFs.
- Lower bounding the value of the on-axis MTF (it is an inequality constraint). This constraint acts somewhat like the minimization of the spot diagram diameters in order to get good image quality.

In practice, it is sufficient to enforce similarity of the MTFs at only two or three well chosen spatial frequencies. This method is very simple to implement in an optimization routine and the modifications of the weights of each constraint can be done in an intuitive manner to drive the whole optimization process on its way.

The conventional triplet is used as a starting point. During the optimization process, we use the same variables and the same general constraints as in the former conventional optimization in Sec. 3 (See Supplemental 1, section 1). We also use the procedure previously explained to switch from fictitious (n,V) glasses to real catalog materials. This results in the triplet shown in Fig. 1(b), whose precise characteristics are given in Supplemental 1 (section 2). It will be referred to as "EMTF triplet". On a structural point of view, this new configuration has a shape similar to the conventional triplet. The main difference is that the central biconcave lens has become larger and its material has changed (NSF8 instead of NSF10).

4.2 Analysis of MTFs and effective MTFs

Figure 2(c) shows the MTFs obtained with the EMTF triplet. Contrary to the very disparate MTFs of the conventional triplet (Fig. 2(a)), these MTF curves are very close to each other. No MTF (in any field position, and neither in tangential nor in sagittal) presents nullings. On the other hand, they are, on average, lower than in the conventional case. For example, the on-axis MTF, which is quite high in the conventional case, takes values below 0.2 at a spatial frequency of 40 lp.mm^{-1}, whereas the Nyquist frequency is 100 lp.mm^{-1}. It goes below the relative noise curve at 60 lp.mm^{-1}. The performance on the axis is therefore lower than for the conventional system.
This difference in optical performance is confirmed by the RMS diameter of the spot diagrams, which are much larger - by a factor of about 5 - in the co-designed case (Fig. 5(b)) than in the conventional case (Fig. 5(a)). The shape of the spot diagrams Fig. 5(b) is characteristic of the presence of strong aperture aberrations (i.e. spherical aberration), which means that the MTF equalization co-optimization method deliberately added aperture aberrations to the original system, most probably to be able to decrease the field-variant aberrations, but also in a special way that does not introduce early MTF nullings.
The study of this EMTF triplet would not be complete without considering the digital processing by the average Wiener filter for which it was (implicitly) co-optimized. Fig.2(d), the effective MTFs obtained after deconvolution are, whatever the field position, close to 1 up to the spatial frequency 40 lp.mm^{-1}. This means that the reconstruction will be possible for all patterns with frequency below 40 lp.mm^{-1}, regardless of their position within the FoV. This was not the case for the conventional triplet, where the effective MTFs (Fig. 2(b)) took very heterogeneous values, reaching zero at some positions in the field. The effective MTFs of the EMTF triplet smoothly decrease only after 40 lp.mm^{-1}, due to the necessary trade-off made by the average Wiener filter between signal and noise.

4.3 Analysis of final image Quality

To compare the performance of the hybrid systems based on optimized triplets, we will use the IQ_{mean} (Eq. 8), Tab. 1. It is seen that without deconvolution, the

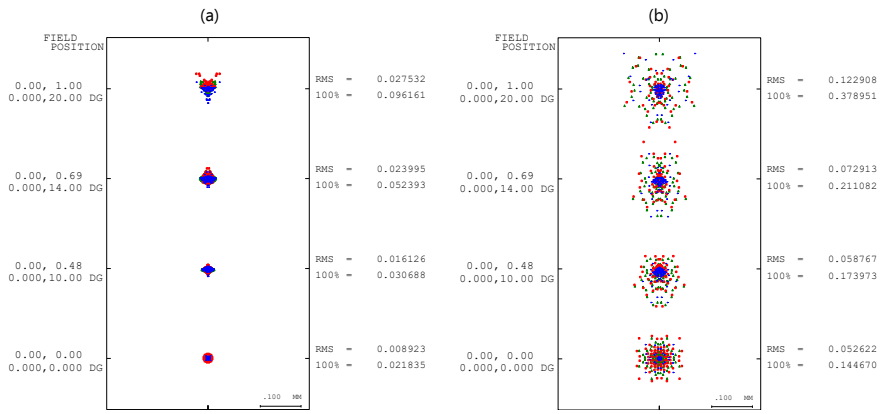

Figure 5: Spot diagrams of (a) the conventional triplet and (b) the EMTF triplet.

EMTF triplet yields a IQ_{mean} 2 dB lower than the conventional triplet. After deconvolution, the performance of both triplets is identical: the performance gain by deconvolution is thus very important for the EMTF triplet (2.6 dB) and exactly compensates for the drop observed before deconvolution. This large performance gain after deconvolution can be easily seen by comparing the first row (without deconvolution) to the second (with deconvolution) of Fig. 6. In the first case, the images appear extremely blurred, much more than in the conventional case (Fig. 6(a-b-c)). On the other hand, the details appear well restored after deconvolution (Fig. 6(d-e-f)).

In order to facilitate the comparison between the hybrid optical/processing systems, two details of the image are precisely analyzed Fig. 7. They belong to the on-axis subpart image and to the extreme field subpart image. The observation of these two details provides a good understanding of how the differences in MTFs play into the imaging performance. We have already observed that the conventional triplet shows an over contrast in the on-axis position (Fig. 7(c)) compared to the ideal scene image (Fig. 7(a)). This over contrast is due to the excessively high value of the on-axis effective MTF at low spatial frequencies. On the other hand, the effective MTFs for larger fields are relatively low, which means that the reconstruction is poor. In particular Fig. 7(d), the pinnacle ornaments are barely visible. Conversely, the MTFs of the EMTF triplet are very close to each other, which explains that imaging performance is similar for the on-axis detail (Fig. 7(e)) and the extreme field detail (Fig. 7(f)): the contrast is less important on axis than for the conventional triplet, but the reconstruction is better at the extremity of the FoV, where the ornaments of the pinnacle are clearly visible.

It can also be noted that with the EMTF triplet, the deconvolved image is slightly more noisy than with the conventional one: this can be seen with the

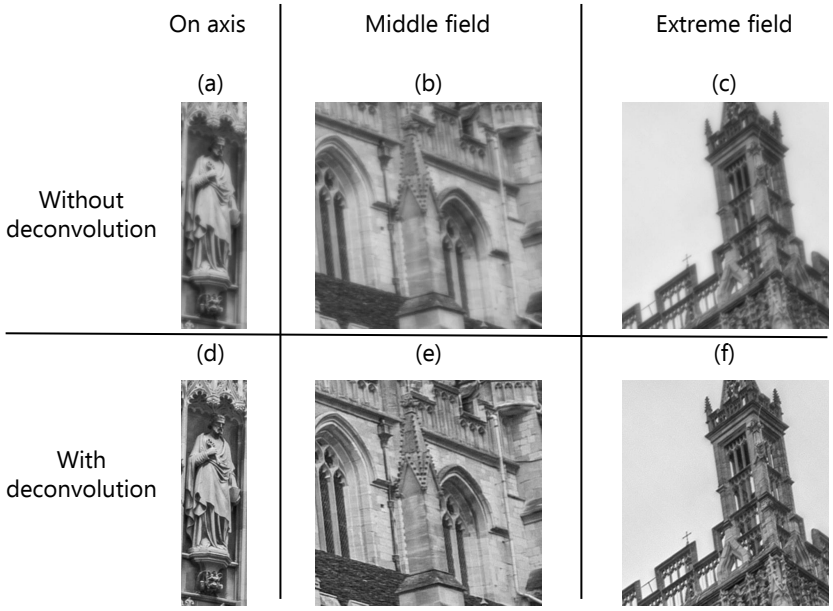

Figure 6: Details of the cathedral image obtained with the EMTF triplet for an SNR of 34 dB, (a-b-c) without deconvolution and (d-e-f) with deconvolution.

slight "orange skin" effect in the sky, Fig. 7(f). This observation suggests an increased sensitivity to noise of the hybrid system based on the EMTF triplet. This was expected since the MTFs of the EMTF triplet are globally lower than those of the conventional one, so that the frequency-wise SNR before deconvolution is smaller. However, since the MTFs of the EMTF triplet remain globally higher than the noise level, deconvolution is still effective, even in the extreme field (Fig. 7(f)): the noise is increased (and at the origin of the "orange skin" effect), but the signal is clearly enhanced. On the contrary, for the hybrid system based on the conventional triplet (Fig. 7(d)), the noise is not increased, so the sky remains smooth, but the signal is not well reconstructed, as the effective MTF is too low.

To conclude, co-optimization of the Cooke triplet by MTF equalization leads to similar average performance as the conventional triplet, after deconvolution. However, the distribution of the performance over the FoV is different: while the conventional method favors the positions in the field close to the axis at the expense of the extreme fields, the EMTF method homogenizes the performance, as originally intended.

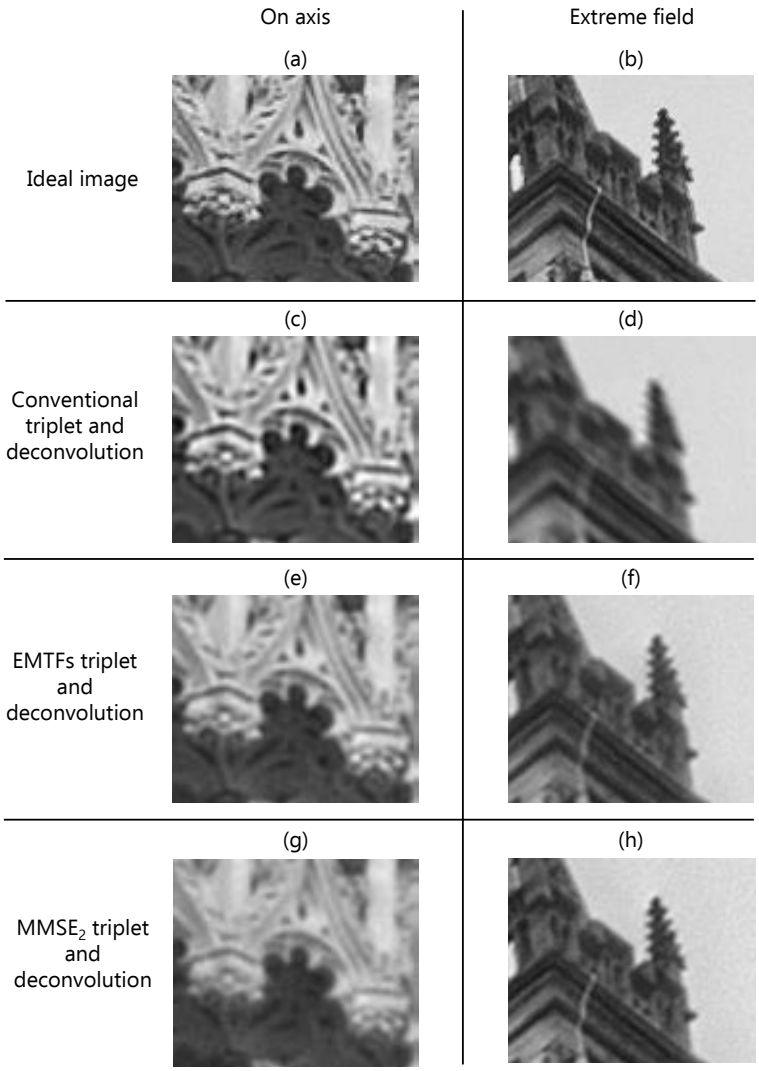

Figure 7: On axis (first column) and the extreme field (second column) details of the cathedral scene. (a-b) Ideal images. (c-d) Conventional triplet and deconvolution. (e-f) EMTF triplet and deconvolution. (g-h) MMSE$_2$ triplet and deconvolution.

5 The MSE as a co-design criterion

5.1 Implementation and use of the MSE criterion

So far, we have used the MSE to evaluate the performance of hybrid systems (through the IQ criterion), but not as a direct co-optimization criterion. An implementation of this criterion has already been done under Zemax [30, 36, 35], but until recently, it seemed to us impossible to do the same under the other well-known CodeV commercial lens design software, since it was not possible to directly access PSFs (or 2D OTFs) in a user-defined error function during optimization. Using the new features of CodeV version $\geq$ 11.0 (2017) that remove these limitations, we have been able to implement – for the first time to our knowledge - MSE optimization directly into CodeV. The details of the implementation are given in Supplemental 1 (section 3). In this section, we co-optimize a Cooke triplet with this method and compare its performance with the two previously described triplets. This allows a fair comparison since all these triplets have been optimized with the the same lens design software.
As for any optimization, the final solution, which corresponds to a local minimum of the criterion, may depend on the chosen starting point. We have thus considered two different starting points. The "MMSE_1 triplet", represented on Fig. 1(c), has been obtained by choosing the conventional triplet as a starting point. The "MMSE_2 triplet", displayed on Fig. 1(d), has been obtained by starting from the EMTF triplet. The detailed characteristics of these two triplets a given in Supplemental 1 (section 2). By comparing Fig. 1(c) and Fig. 1(a) on the one hand, and Fig. 1(b) and Fig. 1(d) on the other hand, it is clear that the MMSE triplets have kept the main structural features of their starting points. However, we will see in the following that its imaging performance is different.

5.2 Analysis of MTFs and effective MTFs

Figure 2(e) represents the MTFs of the MMSE_1 triplet. We can observe that the on-axis MTF collapses quickly and shows a nulling at 50 lp.mm^{-1}. On the other hand, the other MTFs no longer show any nulling for low spatial frequencies, which is a significant improvement over the conventional system. As a result, Fig. 2(f), the effective MTFs higher than 0.38 for spatial frequencies under 40 lp.mm^{-1}, which was not the case in the conventional case (Fig. 2(b)). However, the on-axis effective MTF presents a nulling (the same as for the on-axis MTF). The optimization of MSE_{mean} (see Eq. 6) thus clearly gave priority to the peripheral positions in the FoV over the on-axis MTF, whereas the on-axis MTF is more often favored in conventional optical design because there is naturally less aberration around the optical axis. This evolution was possible because in the MSE-based criterion, the on-axis performance does not have a larger weight than any other field.
Figure 2(g) displays the MTFs of the MMSE_2 triplet. The MTFs are close to each other, as for the EMTF triplet. Similarly, the effective MTFs are close to each other (Fig. 2(h)). The MMSE_2 triplet is the only system that has high

effective MTFs over a wide range of spatial frequencies, with most of its effective MTFs higher than 0.4 at 60 lp.mm^{-1}. In terms of effective MTFs, it can thus be considered as the best solution.

5.3 Analysis of final image quality

Let us now analyze the obtained image quality. It is seen in Table 1 that the MMSE_1 triplet and the MMSE_2 triplet both yield enhanced image quality after deconvolution. This was expected since they are optimized to have a smaller MSE_{mean}, but the important conclusion is that the performance gain is significant. Indeed, for the MMSE_1 triplet, the image quality after deconvolution is increased by 1.1 dB. This means that the optimization on the MSE, by tightening together the previously disparate MTFs (Fig. 2(a) and Fig. 2(e)), made the deconvolution by the average Wiener filter more efficient.
Similarly, the MMSE_2 triplet yields a lower image quality than its starting point (the EMTF triplet) before deconvolution, but a much better one after deconvolution. It shows the best average performance after deconvolution, reaching $IQ_{mean} = 16$ dB. The performance gain with respect to the starting point is even larger in this case, since it amounts to +2.5 dB. Such a large increase in IQ_{mean} can be explained by the fact that the effective MTFs are close to 1 for a wider range of spatial frequencies.
This increased performance of the MMSE_2 triplet is confirmed in the simulated images, Fig. 8. Contrary to the conventional triplet, no over contrast is visible on the restored on-axis subpart image (Fig. 8(d)). The middle field (Fig. 8(e)) and extreme field (Fig. 8(f)) subpart images also appear very well restored. It is possible to observe the reconstruction more in details on Fig. 7(g-h). Compared to the other hybrid systems, we notice that when using MMSE_2 triplet, the restored image is slightly more blurred on the axis, but much sharper in the field: it is the method that best reconstructs the pinnacle ornaments. This behavior was predictable from our analysis of the effective MTFs.
These results lead to several interesting conclusions. The main one is that the hybrid system based on the MMSE_2 triplet has the best performance among the systems considered in this article, both in terms of quantitative and visual image quality. This clearly demonstrates the benefit of implementing MSE-based end-to-end optimization directly in a commercial lens design software. The second conclusion is that both MMSE_1 and MMSE_2 triplets have kept the characteristics of their starting points in terms of optical architecture and properties (heterogeneity or homogeneity of MTFs). Since the MMSE_2 triplet performs better, this shows that, to reach our optimization goal (good image quality at all FoV positions), it was preferable to choose as a starting point a lens having already similar MTFs over the FoV. However, we can also suspect that MSE optimization, as implemented, may have difficulties to "jump out" local minima. The sensitivity of lens design to starting point is a long-standing problem, and a lot of efforts have been done to bring solutions to it in the framework of conventional optical design. An important perspective to the

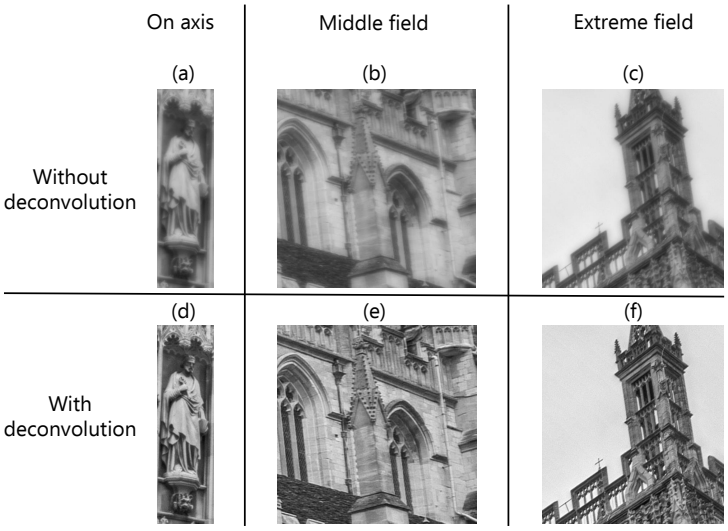

Figure 8: Details of the cathedral image obtained with the MMSE_2 triplet for an SNR of 34 dB, (a-b-c) without deconvolution and (d-e-f) with deconvolution.

present work is to devote the same efforts to enhance the capacity of MSE-based co-optimization to find globally optimal solutions.

6 Conclusion

In this article, we have compared three different methods to co-optimize a hybrid optical/digital imaging system with commercial lens design software: conventional optimization based on minimization of the RMS diameter of spot diagrams, a surrogate criterion based on near equality of the MTFs and a true MSE criterion taking explicitly into account the digital processing in the optimization process. To implement the latter method, we integrated – for the first time to our knowledge – MSE optimization to the CodeV software. These three methods have been illustrated and compared on a concrete application: the design of a Cooke triplet having good image quality everywhere in the FoV. The obtained results demonstrate the superiority of MSE co-optimization over the other methods, both in terms of quantitative and visual image quality.
Analyzing in detail the different resulting lenses, we have shown that the conventional triplet tends to favor the center of the FoV at the expense of the extreme positions, whereas the hybrid system based on the MSE (the MMSE_2 triplet) yields much more homogeneous performance thanks to co-optimization with deconvolution. This homogeneity comes at the expense of the performance on axis, but enable much better performance at peripheral FoV positions. This shows that it is possible, by leveraging deconvolution during the optimization

process, to adapt the spatial distribution of imaging performance to a prescribed goal- which was, here, to reach good performance over the whole FoV.
This work has many perspectives. First, we assumed that the deconvolution algorithm was a unique linear filter common to all FoV positions. This post-processing method is clearly not adapted to the conventional triplet, whose MTFs are quite inhomogeneous over the FoV. They are much more homogeneous with the systems resulting from the two other methods, but there is still residual inhomogeneity. To take into account the inhomogeneity of the MTFs over the FoV and improve further the performance, it would be interesting to consider spatially variant post-processing algorithms [7] in the co-optimization process. Furthermore, recent works have proposed to use home-made differentiable ray tracing connected with a deep neural-network [28, 25, 19, 31, 17, 14]. Comparison with the approach proposed in this paper, that uses a powerful commercial software and a simpler (and faster) post-processing algorithm, would be very interesting. Moreover, it would be interesting to improve the MSE optimization method, by improving for example its speed of execution, its resilience to the choice of the starting point, and also to study the effect of the SNR parameter, or optimize, using the multi-configuration possibility of the lens design program, an optical system that should optimally work for different apertures (e.g. variable iris diaphragm) and therefore different SNRs.

A Supplemental 1

This supplemental document presents the optical constraints used for the co-optimization of the Cooke triplets presented in this article (Sec. A.1), the detailed characteristics of the Cooke triplets co-optimized with various methods (Sec. A.2), and the MSE optimization (Sec. A.3).

A.1 Constraints for optimization of the Cooke triplets

The general constraints applied to the four Cooke triplets are given in Table 2. For the (fictitious) glasses, which can continuously evolve during optimization in the (n, V) or $(n, \Delta n)$ map, where n is the refractive index, V the Abbe number and Δn the blue-red dispersion with $V = (n-1)/\Delta n$, we use the default CodeV constraints that enforce the glass coordinates $(n, \Delta n)$ to stay inside a realistic glass map zone defined by a convex polygon whose vertices correspond to the Schott glasses N-FK5, N-SK16, N-LAF2, SF4 [3, 4, 5].

Applied constraints	CodeV mnemonic	Length [mm]
Minimal center thickness	MNT	1.3
Maximal center thickness	MXT	5
Minimal edge thickness	MNE	1.3
Minimal air center thickness	MNA	0.5
Minimal air edge thickness	MAE	0.5

Table 2: Constraints applied on Cooke triplets (Strict inequality constrains for all relevant lens elements enforced by the Lagrange multipliers method).

A.2 Characteristics of the co-optimized Cooke triplets

We give here the characteristics of the different Cooke triples co-optimized in this article: the conventional triplet (Tab. 3), the EMTF triplet (Tab. 4), the $MMSE_1$ triplet (Tab. 5) and the $MMSE_2$ triplet (Tab. 6). Selected glasses are SCHOTT glasses.

Surface	Radius of curvature [mm]	Thickness [mm]	Glass
Object	∞	∞	—
1	15.99	5.00	N-LAF2
2	34.61	2.72	—
3 (STOP)	-49.47	1.30	N-SF10
4	16.62	2.75	—
5	37.10	3.21	N-LAF2
6	-29.05	41.50	—
Image	∞	-0.20	—

Table 3: Characteristic of the conventional Cooke triplet.

Surface	Radius of curvature [mm]	Thickness [mm]	Glass
Object	∞	∞	—
1	14.39	4.90	N-LAF2
2	27.42	1.50	—
3 (STOP)	-57.93	2.22	N-SF8
4	14.44	2.04	—
5	29.37	3.13	N-LAF2
6	-34.37	41.06	—
Image	∞	-0.18	—

Table 4: Characteristic of the EMTF triplet (Cooke triplet optimized by MTFs equalization).

Surface	Radius of curvature [mm]	Thickness [mm]	Glass
Object	∞	∞	—
1	15.97	4.96	N-LAF2
2	34.58	2.68	—
3 (STOP)	-49.94	1.31	N-SF10
4	16.45	2.73	—
5	36.16	3.40	N-LAF2
6	-29.36	41.58	—
Image	∞	-0.23	—

Table 5: Characteristic of the MMSE_1 triplet (Cooke triplet obtained from the conventional triplet with the MSE criterion).

Surface	Radius of curvature [mm]	Thickness [mm]	Glass
Object	∞	∞	—
1	14.41	4.77	N-LAF2
2	27.76	1.58	—
3 (STOP)	-56.24	2.28	N-SF8
4	14.30	1.98	—
5	28.79	2.99	N-LAF2
6	-34.13	41.14	—
Image	∞	-0.18	—

Table 6: Characteristic of the MMSE_2 triplet (Cooke triplet obtained from the EMTF triplet with the MSE criterion).

A.3 Implementation of MSE-based co-optimization in CodeV

CodeV lens design software offers the ability to write advanced optimization macros, that may use externally coded efficient compiled extension in the form of Dynamic Link Libraries (DLLs). It also offers, since version 11.0, the possibility to calculate the point spread function (PSF) of a given field during the iterations of its internal optimization algorithm, thanks to the predefined CodeV Macro-PLUS function PSF_DATA_1FLD. Therefore, we create a DLL, based on a code written in C, taking as arguments the inputs (2D arrays describing the PSFs) and the output (the MSE) which is calculated according to the values of the PSFs. Note that computing a too large number of PSFs during an optimization routine would not be practicable due to large computing time. Only thirteen of them have been used during the optimization. The distribution of these thirteen fields, used for the construction of the average Wiener filter useful for the deconvolution of the image obtained with a Cooke triplet of half-field 20°, is presented in Fig. 9. Their positions guarantee a nearly rotational symmetry of the average Wiener filter, so that it has no preferred direction.

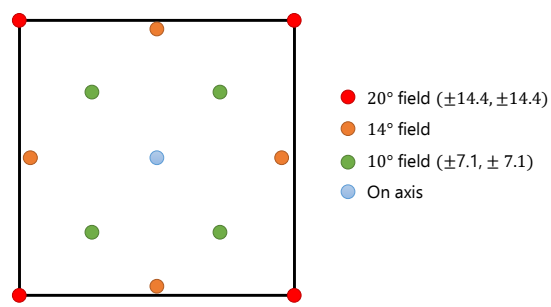

Figure 9: Representative scheme of the $K = 13$ positions in the field considered for the construction of the average Wiener filter useful for the deconvolution of the image obtained with an optical system of half-field 20°.

In order to compute the value of $MSE_{mean}(\Psi)$ (with here $K = 13$) using Eq.(3) of the main paper that expresses it in the Fourier domain, it is useful to use the *FFTW* free library available in C, which allows to efficiently compute the Discrete Fourier Transform of the PSFs, and thus obtain the optical transfer functions $\tilde{h}_{\psi_k}$. This implemented MSE does not take into account defects that do not directly affect the shape of PSFs such as relative intensity change and distortion (which can be otherwise constrained if necessary during optimization). For optimization purposes, these two defects will be considered negligible or easily corrected by post-processing, as they generally have little effect on the visual quality of an image. The written C code must be compiled as a 32-bits DLL (CodeV is only available for this software architecture) using a Synopsys-supplied Makefile, extended to allow *FFTW* use. This allows one to obtain a DLL (and associated files) that can easily be called in a homemade CodeV

macro function. This function can then be called directly in the CodeV optimization routine, like other functions of interest to the optical designer. The optimization goal is then to make the obtained MSE result as close as possible to 0.

Acknowledgments The work reported in this study is supported in part by the Agence de l'Innovation de Défense (AID) that provides half of a PhD fellowship to Alice Fontbonne.

Disclosures The authors declare no conflicts of interest.

Data Availability Statement Data is available upon request.

References

[1] Marie-Anne Burcklen, Hervé Sauer, Frédéric Diaz, and François Goudail. Joint digital-optical design of complex lenses using a surrogate image quality criterion adapted to commercial optical design software. *Applied Optics*, 57(30):9005, oct 2018.

[2] W. Thomas Cathey and Edward R. Dowski. New paradigm for imaging systems. *Appl. Opt.*, 41(29):6080–6092, Oct 2002.

[3] Code V. *CODE V Introductory User's Guide*. Synopsys Code V, 2021.

[4] Code V. *CODE V Lens System Setup Reference Manual*. Synopsys Code V, 2021.

[5] Code V. *CODE V Optimization Reference Manual*. Synopsys Code V, 2021.

[6] Geoffroi Côté, Jean-François Lalonde, and Simon Thibault. Deep learning-enabled framework for automatic lens design starting point generation. *Optics Express*, 29(3):3841, 2021.

[7] Loïc Denis, Eric Thiébaut, Ferréol Soulez, Jean-Marie Becker, and Rahul Mourya. Fast approximations of shift-variant blur. *Int. J. Comput. Vision*, 115(3):253–278, December 2015.

[8] Frédéric Diaz, François Goudail, Brigitte Loiseaux, and Jean-Pierre Huignard. Increase in depth of field taking into account deconvolution by optimization of pupil mask. *Optics letters*, 34(19):2970–2972, 2009.

[9] Donald C. Dilworth. Pseudo-second-derivative matrix and its application to automatic lens design. *Applied Optics*, 17(21):3372, 1978.

[10] Edward R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34(11):1859–1866, Apr 1995.

[11] Rafael Falcón, François Goudail, Caroline Kulcsár, and Hervé Sauer. Performance limits of binary annular phase masks codesigned for depth-of-field extension. *Optical Engineering*, 56(6):065104, 2017.

[12] Robert Fisher, Biljana Tadic-Galeb, and Paul Yoder. *Optical System Design, Second Edition*. McGraw-Hill, 2008.

[13] Alice Fontbonne, Hervé Sauer, and François Goudail. Theoretical and experimental analysis of co-designed binary phase masks for enhancing the depth of field of panchromatic cameras. *Optical Engineering*, 60(03), mar 2021.

[14] A. Halé, P. Trouvé-Peloux, and J.-B. Volatier. End-to-end sensor and neural network design using differential ray tracing. *Optics Express*, 29(21):34748, 2021.

[15] Thomas Houllier and Thierry Lépine. Comparing optimization algorithms for conventional and freeform optical design. *Optics Express*, 27(13):18940, 2019.

[16] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2(2):164–168, 1944.

[17] Yuankun Liu, Chongyang Zhang, Tingdong Kou, Yueyang Li, and Junfei Shen. End-to-end computational optics with a singlet lens for large depth-of-field imaging. *Optics Express*, 29(18):28530, 2021.

[18] Donald W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[19] Elias Nehme, Boris Ferdman, Lucien E. Weiss, Tal Naor, Daniel Freedman, Tomer Michaeli, and Yoav Shechtman. Learning optimal wavefront shaping for multi-channel imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2179–2192, 2021.

[20] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C (2nd Ed.): The Art of Scientific Computing*. Cambridge University, New York, NY, USA, 1992.

[21] Dirk Robinson and David G. Stork. Joint design of lens systems and digital image processing. In *International Optical Design*, page WB4. Optical Society of America, 2006.

[22] John R. Rogers. A new method for the optimization of optical systems: Comparisons and discussions. In *Journée Thématique Calcul Optique, Société Française d'Optique*, jan 2014.

[23] Daniel L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385 – 3398, 1997.

[24] Furkan E. Sahin. Open-source optimization algorithms for optical design. *Optik*, 178:1016–1022, 2019.

[25] Ruibo Shang, Kevin Hoffer-Hawlik, Fei Wang, Guohai Situ, and Geoffrey P. Luke. Two-step training deep learning framework for computational imaging without physics priors. *Optics Express*, 29(10):15239, 2021.

[26] K. D. Sharma. Design of a new five-element cooke triplet derivative. *Applied Optics*, 18(23):3933, 1979.

[27] K. D. Sharma. Four-element lens systems of the cooke triplet family: designs. *Applied Optics*, 19(5):698, 1980.

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

[29] Tim Stangner, Tobias Dahlberg, Pontus Svenmarker, Johan Zakrisson, Krister Wiklund, Lene B. Oddershede, and Magnus Andersson. Cooke–triplet tweezers: more compact, robust, and efficient optical tweezers. *Optics Letters*, 43(9):1990, 2018.

[30] David G. Stork and M. Dirk Robinson. Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems. *Appl. Opt.*, 47(10):B64–B75, Apr 2008.

[31] Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. End-to-end complex lens design with differentiable ray tracing. *ACM Trans. Graph.*, 40(4), 2021.

[32] H Dennis Taylor. Optical designing as an art. *Transactions of the Optical Society*, 24(3):143, 1923.

[33] A. van der Schaaf and J.H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17):2759 – 2770, 1996.

[34] Darko M. Vasiljevic. Optimization of the cooke triplet with various evolution strategies and damped least squares. In Richard C. Juergens, editor, *Optical Design and Analysis Software*. SPIE, 1999.

[35] Tom Vettenburg and Andrew R. Harvey. Holistic optical-digital hybrid-imaging design:wide-field reflective imaging. *Appl. Opt.*, 52(17):3931–3936, Jun 2013.

[36] Jiaoyang Wang, Lin Wang, Ying Yang, Rui Gong, Xiaopeng Shao, Chao Liang, and Jun Xu. An integral design strategy combining optical system and image processing to obtain high resolution images. In Bormin

Huang, Chein-I Chang, and Chulhee Lee, editors, *Remotely Sensed Data Compression, Communications, and Processing XII*. SPIE, 2016.