# Comparison of three methods for end-to-end optimization of hybrid optical/digital imaging systems with professional optical design software

Alice Fontbonne, Hervé Sauer, François Goudail

# Comparison of three methods for end-to-end optimization of hybrid optical/digital imaging systems with professional optical 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 different methods to optimize end-to-end a hybrid optical/digital system for best and most uniform performance over the field-of-view with the Synopsys® CodeV® lens design software. We have extended the native optimization capability of this software by implementing different methods that leverage the deconvolution during the optimization process of the hybrid optical system as a whole, including simultaneously their optical and digital image processing parts. We show that the joint optimization of the lens and the processing through a true restored-image quality criterion significantly enhances the final post-processed image quality, and allows to fine-tune the residual balancing between on-axis and peripheral fields with simple weighting coefficients.

**keywords:** Complex lens; end-to-end optimization; hybrid system; optical system design

*This author version has been published as:* Alice Fontbonne, Hervé Sauer, and François Goudail, "Comparison of three methods for end-to-end optimization of hybrid optical/digital imaging systems with professional optical design software", Proc. SPIE 12136, Unconventional Optical Imaging III, 121360O (20 May 2022)
DOI : https://doi.org/10.1117/12.2621453

## 1 INTRODUCTION

Many imaging systems consist today of a complex optical system followed by a digital processing algorithm. It seems therefore interesting to optimize the whole imaging chain - including post-processing - instead of separating the optimization of the optical part (i.e, the lens) from that of the processing part, as is still regularly done. This approach is called "optical/digital co-design" or "end-to-end optimization". Two different approaches are currently investigated to implement this approach. The first one consists in developing a ray tracing software that computes both the image provided by the lens and its derivatives with respect to the lens parameters. This information is then sent to a neural network (NN) that processes the image. Both the lens and the NN parameters are optimized with a stochastic gradient algorithm [5, 8]. The objective of this method is to be totally automatic and to dispense with the experience of the optical designer. We chose an alternative approach, which consists in taking advantage of commercial lens design software (like Zemax® OpticStudio®, Synopsys® CodeV®, etc.), whose ray tracing and optimization algorithm qualities are well known, and which constitute a familiar environment for the lens designer. These software packages do not natively take into account the influence of post-processing algorithms on lens design, but this capability can be added to them in order to implement end-to-end optimization [7, 10, 11, 1].

In this paper, we investigate and compare three methods for optimizing a hybrid system end-to-end with CodeV: a conventional method (as a reference) that does not take digital processing

into account at the optimization stage, a "mixed method" that *implicitly* takes it into account, and real "end-to-end" method that explicitly minimizes the mean square error (MSE) between the processed image and a "ground truth". Considering, as a use-case example, optimization of a Cooke triplet to obtain the best and most uniform image quality in the field-of-view (FoV), we show that the MSE-based method is the most efficient one, and that the MSE criterion may be tuned with simple weighting coefficients to change at will the balancing between the residual image quality discrepancy between on-axis and peripheral fields. In Sec. 2, we describe the details of these optimization methods. Then, in Sec. 3, we apply them to the optimization of a Cooke triplet with wide FoV and compare the obtained image performance on the basis of quantitative quality metrics and qualitative analysis of simulated images.

# 2 OPTIMIZATION METHODS

Let us consider the set $\Psi = \{\psi_1, \psi_2, ..., \psi_K\}$ of $K$ field positions spanning the whole FoV. We denote $\tilde{h}_{\psi_k}(\nu)$ the optical transfer function (OTF) of the imaging system at the position $\psi_k$, $\nu$ the spatial frequency, $S_{oo}(\nu)$ the statistical power spectral density (PSD) of the scene (with $\int S_{oo}(\nu)d\nu = 1$) and $S_{nn}(\nu)$ the PSD of the noise (set so that the signal-to-noise ratio on the raw acquired image is 34 dB, with $SNR = 10\log_{10}\left[\int S_{oo}(\nu)d\nu / \int S_{nn}(\nu)d\nu\right]$). For $S_{oo}(\nu)$, as natural scenes are well represented with a power-law PSD [6, 9], we use a generic ideal image model $S_{oo}(\nu) \propto \nu^{-2.5}$. The MSE between the final (restored) image and the ground truth at a position $\psi_k$, averaged over image and noise realizations, can be expressed in the Fourier domain as [2]:

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

By averaging the MSE over all the field positions, one obtains the *average* MSE defined as :

$$MSE_{mean}(\Psi) = \frac{1}{K}\sum_{k=1}^{K} MSE(\psi_k) \tag{2}$$

The *average* Wiener filter is the linear deconvolution filter that minimizes $MSE_{mean}$ [2]:

$$\tilde{w}_{\Psi}(\nu) = \frac{\frac{1}{K}\sum_{k=1}^{K} \tilde{h}_{\psi_k}(\nu)^{\star}}{\frac{1}{K}\sum_{k=1}^{K} |\tilde{h}_{\psi_k}(\nu)|^2 + \frac{S_{nn}(\nu)}{S_{oo}(\nu)}} \quad , \tag{3}$$

In this article, we will consider hybrid systems whose optical part is a Cooke Triplet lens with large FoV (the general specifications of the aimed Cooke triplets are given in Tab. 1) and the processing algorithm is the average Wiener deconvolution filter defined Eq. 3. In addition, we will consider three methods for optimizing a hybrid system:

- The "conventional" method is natively implemented in CodeV. It consists in minimizing the quadratic sum of the spot-diagram diameters over a limited set of field positions laying in one meridional plane. This classical method assumes equal weights for the squared spot-diagram diameters at all the considered field positions, which generally favors the on-axis performance. The lens optimized with this method will be called "CCT" (for Conventional Cooke Triplet).

- The "MTF equalization" method, which implicitly takes into account that the image will be processed with the average Wiener filter. Since it is obvious that for this filter to be efficient, the MTFs at all field positions must be similar to each other and without nulling in the spatial frequency range of interest, we use a "surrogate" optimization criterion that favors equality between the MTF on-axis and the MTFs at a set of field positions, and enforces a lower bound on the value of the on-axis MTF to make deconvolution efficient. This method allows to impose a certain homogeneity of performance across the FoV [1]. The lens optimized with this method will be called "HCT" (for Homogeneous Cooke Triplet).

- The "MSE optimization" takes the image post-processing explicitly into account by minimizing the average MSE defined in Eq. 2. Moreover, we will see in the following that introducing weights for the fields $\psi_k$ makes it possible to achieve controlled balancing between on-axis image quality and off-axis image quality.

| | |
|---|---|
| Focal length | 50 mm |
| Aperture | F/4 |
| FoV | 40 deg ($\pm$20 deg) |

Table 1: Optical specifications of the Cooke triplets.

The MTFs of the CCT at different positions in the field, for tangential (T) and radial (R) orientations, are displayed on Fig. 1(a). As it is common for conventional optical systems (with usual weighting of fields), the on-axis MTF is much higher than the other MTFs. In particular, the MTFs at maximal field positions present quasi-nullings at frequency 40 lp.mm$^{-1}$. Conversely, the MTFs of the HCT are quite close to each other (Fig. 1(b)). This homogeneity of the MTFs leads to reduce the average MSE of the hybrid system when post-processing with the average Wiener filter. On the other hand, the fact that the MTFs are significantly lower than the on-axis MTF of the CCT has an adverse effect on the MSE. Moreover, we can see in Fig. 1(b) that the MTF is the lowest for the on-axis field, which is not, in general, a preferred configuration. In the next section, we investigate how direct optimization of the MSE criterion, which we have implemented in the CodeV software (through a compiled extension written in the C language), allows to further increase the system performance [3, 4]. It will also be shown that some slight modifications of the average MSE expression to introduce weightings between fields, allow to fine-tune the balancing between image quality on-axis and at peripheral fields.

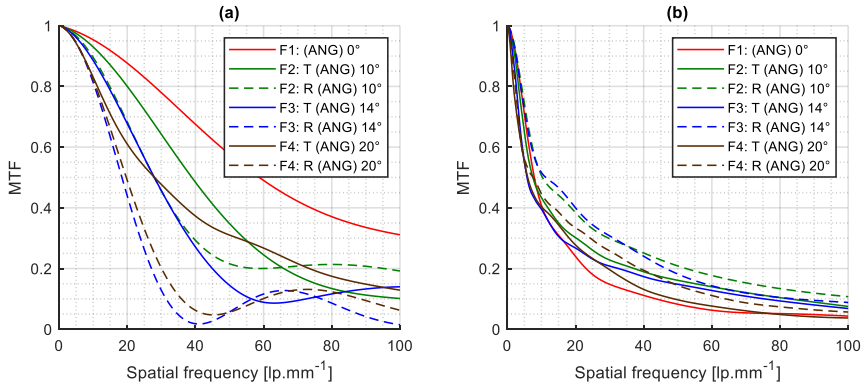


Figure 1: (a) MTF of the CCT (b) MTF of the HCT.

# 3 End-to-end optimization with weighted average MSE criterion

In order to give the optical designer some freedom to emphasize some parts of the field-of-view when it is relevant for a given application, we assign a weight $\alpha_k$ to each field $\psi_k$, and define

the *weighted average* MSE as follows:

$$MSE_p(\Psi) = \frac{1}{A}\sum_{k=1}^{K} \alpha_k MSE(\psi_k) \tag{4}$$

where $A = \sum_{k=1}^{K} \alpha_k$. This results in the following expression for the weighted average Wiener filter which will be used to optimally minimize the weighted average MSE:

$$\tilde{w}_\Psi(\nu) = \frac{\frac{1}{A}\sum_{k=1}^{K} \alpha_k \tilde{h}_{\psi_k}(\nu)^\star}{\frac{1}{A}\sum_{k=1}^{K} \alpha_k |\tilde{h}_{\psi_k}(\nu)|^2 + \frac{S_{nn}(\nu)}{S_{oo}(\nu)}} \quad . \tag{5}$$

The set $\Psi = \{\psi_1, \psi_2, \ldots, \psi_{13}\}$ is made of 13 fields well distributed on the sensor, where $\psi_1$ corresponds to the on-axis position, and is associated to the weight $\alpha_1$. Throughout this article, the value of $\alpha_1 = p$ will vary, and all the other weights $\alpha_k$ (for $k \neq 1$) stay equal to 1. To investigate the efficiency of the MSE-based optimization criterion and the effect of tuning the on-axis weight $\alpha_1$, we analyze in this section systems optimized with different values of $p$. Namely, we consider the criteria $MSE_1$, $MSE_2$, $MSE_3$ and $MSE_4$, constructed respectively with $\alpha_1 = 2$, $\alpha_1 = 3$ and $\alpha_1 = 4$, i.e. multiplying the weight of the on-axis position by 2, 3, and 4, respectively. More precisely, the "MSE1 triplet" is obtained by minimizing the $MSE_1$ criterion with Code V, and using the HCT as the starting point of the optimization. Then, we obtain the "MSE2 triplet" by using the $MSE_2$ criterion and taking the MSE1 triplet as the starting point, then the "MSE3 triplet" by using the $MSE_3$ criterion and the MSE2 triplet as the starting point, and finally the "MSE4 triplet" by using the $MSE_4$ criterion and the MSE3 triplet as the starting point. In the following, we will evaluate the impact of this change of on-axis weight with various quality metrics.

### 3.1 MTFs and effective MTFs

The MTFs of these four MSE triplets are presented on the first column of Fig. 2. We can notice a common feature to the four graphs: the MTFs are close to each other and without any nulling, even though they are somewhat low. They share this characteristic with the HCT, that served as a starting point for the optimization of the "MSE1 triplet". Moreover, for both the HCT and the MSE1 triplet, it is the on-axis MTF that is lower than the others. Nevertheless, increasing the weight of the on-axis field position is effective in gradually rising the on-axis MTF (red curve). For instance, the on-axis MTF value at 20 lp.mm$^{-1}$ is 0.21 for the MSE1 triplet, 0.27 for the MSE2 triplet, 0.28 for the MSE3 triplet and 0.31 for the MSE4 triplet. Interestingly, this improvement does not noticeably decrease the MTF values at peripheral field positions.

Let us now observe on the right column of Fig. 2, the *effective* MTFs of the MSE triplets. The effective MTF is the product of the MTF (at a given position in the FoV) with the frequency response of the deconvolution filter - which is, in our case, the average Wiener filter (Eq. 3, without weighting). In the ideal case, and in the absence of noise, the effective MTF would be equal to 1 at all spatial frequencies. A value greater than 1 at a given spatial frequency indicates over-contrast, while a value smaller than 1 indicates under-contrast relatively to the ideally restored image, which is the ideal image of the scene without any blur due to lens aberrations and diffraction. The effective MTF obtained with the average Wiener filter is not uniformly equal to 1, since it corresponds to the best compromise (in the $MSE_p$ sense) between good reconstruction of the useful part of the signal and mitigation of noise reinforcement. We can notice on Fig. 2(b) that in terms of effective MTF, the MSE1 triplet clearly performs poorer on-axis than in the peripheral positions of the FoV. With the MSE2 triplet, the on-axis curve rises strongly (Fig. 2(d)). Such an improvement is also visible between the MSE2 and the MSE3 triplets. For example, the effective MTF rise from 0.58 at 40 lp.mm$^{-1}$ for the MSE2 triplet (Fig. 2(d)) up to 0.62 for the MSE3 triplet (Fig. 2(f)). Finally, the MSE4 criterion further improves the performance on the axis, keeping the effective MTF higher than 0.7 for spatial frequencies even larger than 40 lp.mm$^{-1}$. Therefore, rising the on-axis weight $\alpha_1$

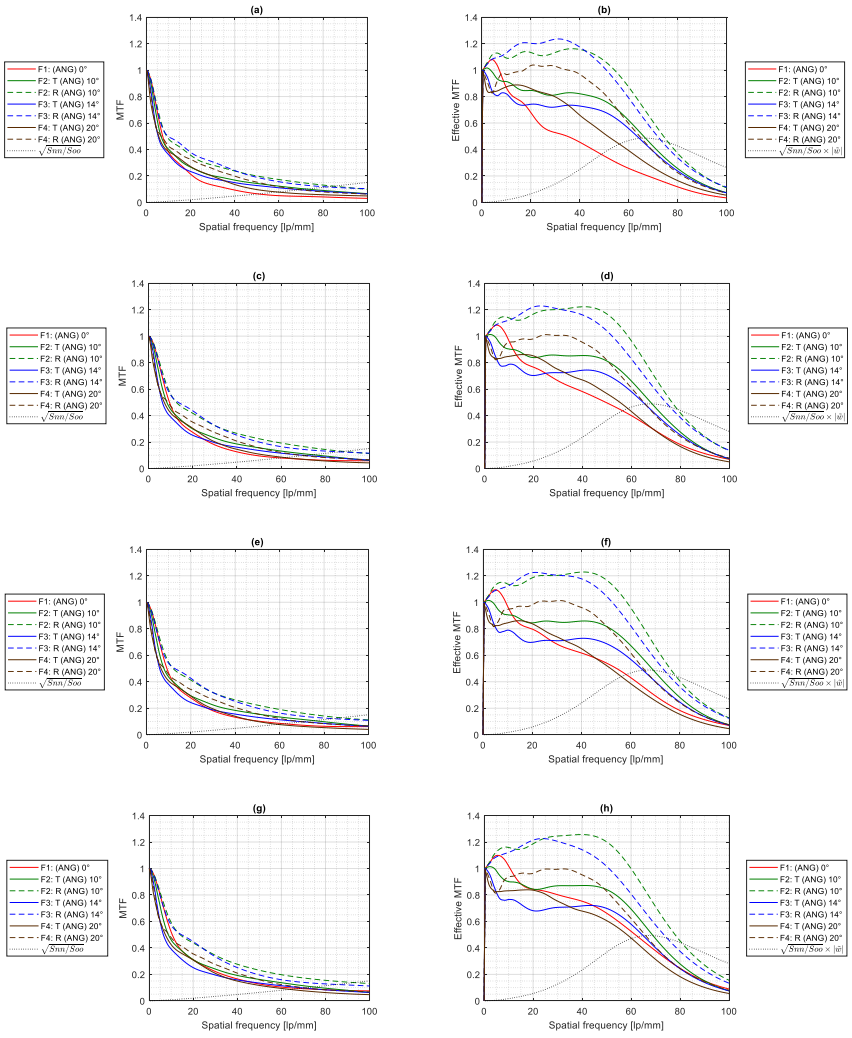


Figure 2: (First column) MTFs (Second column) Effective MTFs, for (a-b) the MSE1 Cooke Triplet, (c-d) the MSE2 Cooke Triplet (e-f) the MSE3 Cooke Triplet and (g-h) the MSE4 Cooke Triplet.

yields a constant improvement of the on-axis MTF and effective MTF without any significant degradation at peripheral positions in the FoV.

## 3.2 Quantitative performance evaluation

The effective MTFs are not sufficient to fully describe the performance of a hybrid system because they do not show the noise reinforcement introduced by the deconvolution filter. A better way to understand the influence of the processing on the final image quality is to evaluate the local MSE at a specific position $\psi_k$ in the FoV. Let us first consider the MSE without deconvolution (which is obtained by substituting, in Eq. 1, $\tilde{w}_\Psi(\nu)$ with a function uniformly equal to 1). Figure 3(a) represents its evolution as a function of the position in the FoV for the different Cooke triplets considered in the article. We notice that the best MSE (without deconvolution) is obtained with the CCT, in particular on the axis, where its value is 8 times

lower than that of the MSE1 triplet. This was expected since the optimization of this triplet does not take digital processing into account. The MSE values obtained with the MSE triplets and the HCT are thus larger and close to each other, which was also expected since they are optimized to work with a digital processing. Moreover, we can notice that the MSE variations are correlated with the variations of the MTF: if the MTF at a given position in the FoV has higher values than another (especially for low spatial frequencies), then the MSE a this position is lower. Looking at the details, we also observe that increasing the weight of the central field (i.e., passing progressively from $MSE_1$ to $MSE_4$) allows to slightly improve the average performance before deconvolution, and more particularly on the axis.

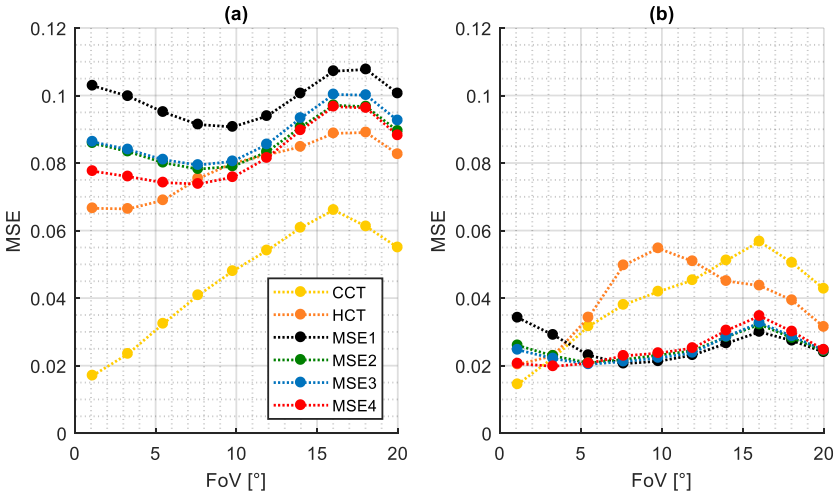


Figure 3: Local MSE as a function of the field position in the FoV (a) without deconvolution; (b) with deconvolution with the average Wiener filter (without weighting).

Figure 3(b) displays the MSE obtained by the different triplets after deconvolution with the average Wiener filter. We can notice that the improvement brought by deconvolution clearly depends on the lens on which it is used. It is quite low for CCT, substantial for HCT and very significant for the different MSE triplets. It is important to note that, as desired, the performance at the center of the FoV gets better and better when increasing the weight on it, i.e., when passing from $MSE_1$ to $MSE_4$. On the other hand, the performance at intermediate positions in the FoV (from 5° on) is slightly degraded, but the four MSE triplets have equal performance at the maximum field position (20°).

This slight degradation of the local MSE with the field angle results in a decrease in average performance (after deconvolution) of the MSE2, MSE3 and MSE4 triplets. The "Image quality" (IQ) defined as:

$$IQ = 10 \times \log_{10}\left(\frac{1}{MSE_{mean}(\Psi)}\right) \quad , \tag{6}$$

where $MSE_{mean}(\Psi)$ is defined in Eq. 2, can be used to quantitatively study performances over the FoV. It is expressed in dB and its value for the 4 considered MSE triplets is given in Tab. 3.2. The MSE1 triplet is, as expected, the best one (16 dB), as it has been optimized with the same criterion that we use to evaluate its performance. However, this performance is not so far from that of the MSE4 triplet (−0.3 dB), which, on the other hand, has a better performance on the axis (Fig. 2(g-h) and Fig. 3(b)). Moreover, the performance with deconvolution of all the MSE triplets is significantly better than that of the CCT and HCT, which is equal to 13.5 dB for both lenses.

| Cooke Triplet | Without deconvolution | With deconvolution with an average Wiener filter |
|---|---|---|
| MSE1 | 10.1 dB | 16.0 dB |
| MSE2 | 10.6 dB | 15.9 dB |
| MSE3 | 10.5 dB | 15.9 dB |
| MSE4 | 10.8 dB | 15.7 dB |

Table 2: IQ of Cooke triplets with or without deconvolution.

### 3.3 Qualitative evaluation on simulated images

A complementary method to evaluate the benefit of weighting the optimization criterion to achieve better quality on-axis is to use image simulations. Figure 4 represents a subpart (close to the on-axis position) of the images obtained with the different MSE triplets. The images in the first column are obtained without deconvolution, and the second column is obtained with deconvolution (Fig. 4(i) being the ground truth, i.e. the ideal scene image). Without deconvolution, we can see the improvement of on-axis visual quality as the weight of on-axis position increases, since from Fig. 4(a) to Fig. 4(g), images are more and more contrasted. Nevertheless, to approach the sharpness of the ground truth (Fig. 4(i)), it is necessary to deconvolve (Fig. 4(second column)). After deconvolution, the visual quality is quite similar between the four MSE triplets, but we can still observe the expected improvement. In particular, the "stone lace work" in the dark background (bottom left part of the image) is slightly blurrier in Fig. 4(b), obtained with the MSE1 triplet, than in Fig. 4(f), obtained with the MSE3 triplet, and in Fig. 4(h), obtained with the MSE4 triplet. This fits well the theoretical conclusions drawn from Fig. 3(b).

Figure 5 represents the same data as in Fig. 4, for another subpart of the image located at a large field angle. We can notice once again the great improvement between the case without deconvolution (Fig. 5(first column)) and the case with deconvolution (Fig. 5(second column)). Moreover, we can observe that after deconvolution, there is very little visual difference between the four MSE triplets, which corroborates the result obtained Fig. 3 for positions in the field between 15° and 20°. To conclude, image simulations leads to similar conclusions than the theoretical predictions and show that optimization with weighted MSE criteria ($MSE_2$, $MSE_3$ and $MSE_4$) can improve the on-axis performance without significantly decreasing the average performance over the whole FoV, and in particular the performance at the peripheral positions of the FoV.

## 4 CONCLUSION

We have compared three different methods to optimize hybrid systems based on a Cooke triplet and a digital deconvolution: A conventional one, as a reference, where the lens is optimized without taking into account the post-processing, which is introduced in a second step; a second one with optimization using MTF equalization criteria that implicitly take into account the image restoration by deconvolution; and a third one that truly optimizes simultaneously the lens and the deconvolution algorithm through a restored-image quality criterion (the MSE), which has been implemented in practice in the CodeV software. For the aim of getting the best and most uniform post-processed image quality all over the field-of-view of a Cooke triplet, we have shown that the second method performs better than the first, and that the third overperforms the second. Moreover, the MSE criterion can be simply tuned with weights in order to emphasize some parts of the FoV, allowing in particular to balance at will image quality in central and peripheral field positions. These results show that a true hybrid system co-optimization criterion can be efficiently implemented in a professional lens design software, while giving the optical designer the flexibility to adapt the optimization criterion to its applications and to use all the software classic tools for system analysis and enforcement of manufacturing constraints.

This work has several perspectives. For now, the MSE triplets remain quite close from the starting point of the optimization. It would be useful to improve the co-design tool using the

weighted MSE criterion in order to better explore the optimization landscape and move farther away from starting point. It is also possible to envisage the implementation of other co-design criteria or other deconvolution algorithms in the lens design software CodeV.

# 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] 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.

[3] Alice Fontbonne. *Conception conjointe combinaison optique / traitement : Une nouvelle approche de la conception optique de haut niveau*. PhD thesis, Université Paris Saclay, 2021.

[4] 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(8):13556–13571, Apr 2022.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

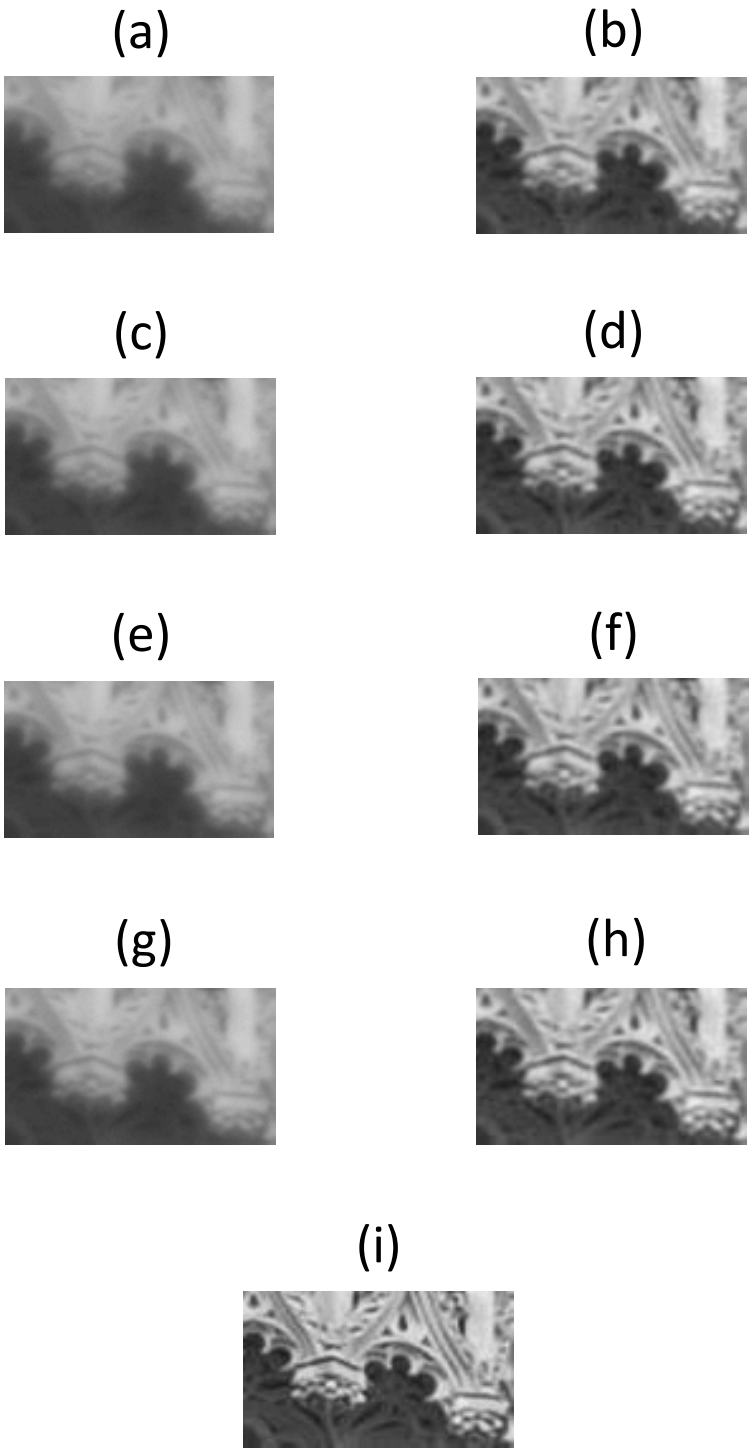


Figure 4: Center part of an image simulated through a hybrid system, (First column) without deconvolution and (Second column) with deconvolution with an average Wiener filter. (a-b) MSE1 Cooke Triplet, (c-d) MSE2 Cooke Triplet (e-f) MSE3 Cooke Triplet and (g-h) MSE4 Cooke Triplet. (i) Ground truth. All simulations of deconvolution have been done with a 34 dB SNR.

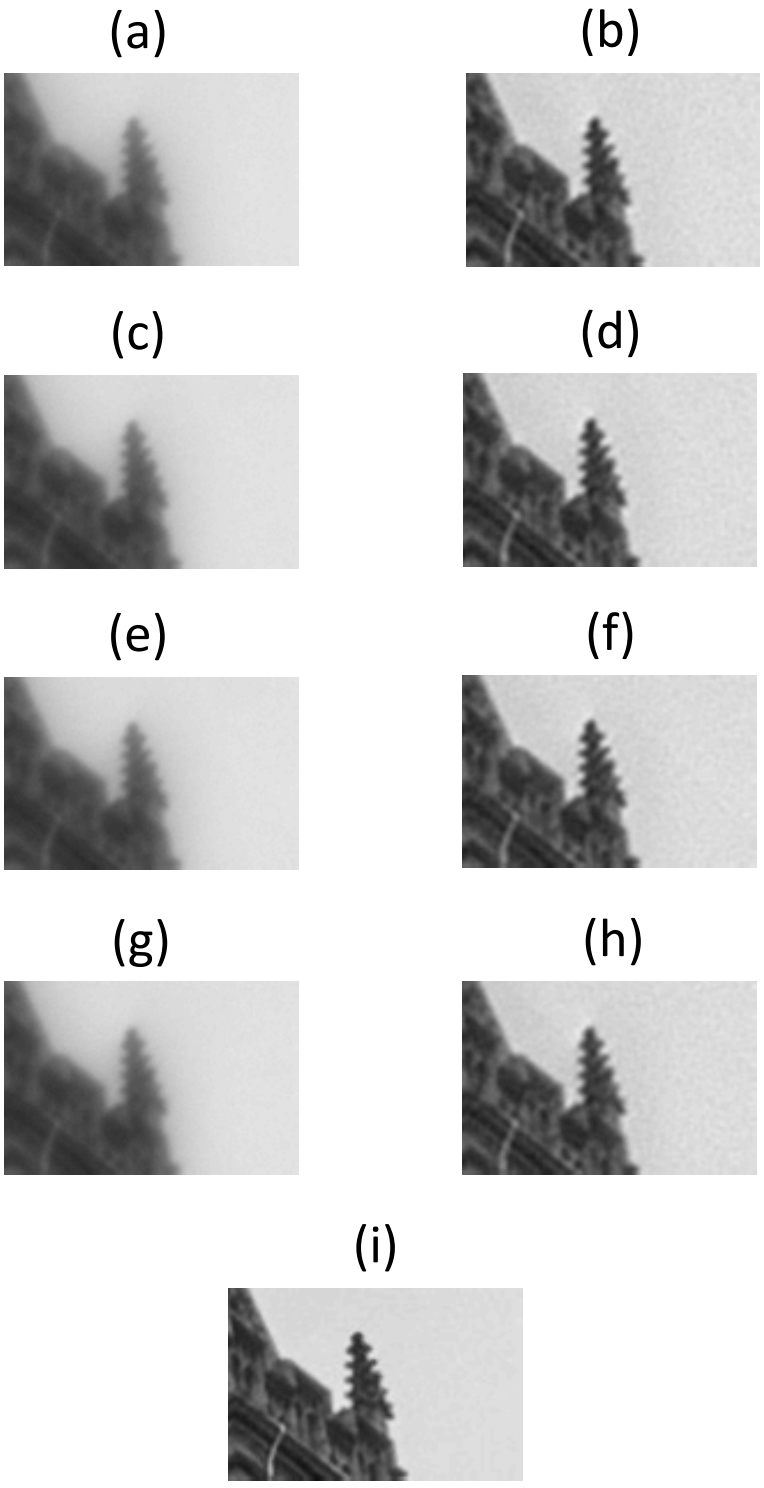


Figure 5: Edge part of an image simulated through a hybrid system, (First column) without deconvolution and (Second column) with deconvolution with an average Wiener filter. (a-b) MSE1 Cooke Triplet, (c-d) MSE2 Cooke Triplet (e-f) MSE3 Cooke Triplet and (g-h) MSE4 Cooke Triplet. (i) Ground truth. All simulations of deconvolution have been done with a 34 dB SNR.