

Experimental demonstration of extended depth-of-field $f/1.2$ visible High Definition camera with jointly optimized phase mask and real-time digital processing

M.-A. Burcklen

marie-anne.burcklen@fr.thalesgroup.com

Thales Angénieux, Boulevard Ravel de Malval, 42570 Saint-Héand, France

F. Diaz

Thales Angénieux, Boulevard Ravel de Malval, 42570 Saint-Héand, France

F. Leprêtre

Thales Angénieux, Boulevard Ravel de Malval, 42570 Saint-Héand, France

J. Rollin

Thales Angénieux, Boulevard Ravel de Malval, 42570 Saint-Héand, France

A. Delboulbé

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

M.-S. L. Lee

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

B. Loiseaux

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

A. Koudoli

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

S. Denel

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

P. Millet

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

F. Duhem

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

F. Lemonnier

Thales Research & Technology France, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France

H. Sauer

Laboratoire Charles Fabry de l'Institut d'Optique, CNRS, Université Paris-Sud, Campus Polytechnique, RD 128, 91127 Palaiseau Cedex, France

F. Goudail

Laboratoire Charles Fabry de l'Institut d'Optique, CNRS, Université Paris-Sud, Campus Polytechnique, RD 128, 91127 Palaiseau Cedex, France

Increasing the depth of field (DOF) of compact visible high resolution cameras while maintaining high imaging performance in the DOF range is crucial for such applications as night vision goggles or industrial inspection. In this paper, we present the end-to-end design and experimental validation of an extended depth-of-field visible High Definition camera with a very small f -number, combining a six-ring pyramidal phase mask in the aperture stop of the lens with a digital deconvolution. The phase mask and the deconvolution algorithm are jointly optimized during the design step so as to maximize the quality of the deconvolved image over the DOF range. The deconvolution processing is implemented in real-time on a Field-Programmable Gate Array and we show that it requires very low power consumption. By mean of MTF measurements and imaging experiments we experimentally characterize the performance of both cameras with and without phase mask and thereby demonstrate a significant increase in depth of field of a factor 2.5, as it was expected in the design step.

[DOI: <http://dx.doi.org/10.2971/jeos.2015.15046>]

Keywords: Wavefront encoding, imaging systems, image processing

1 INTRODUCTION

Most imaging systems now include digital post-processing to enhance image quality and correct defects of the optical system. In classical design approaches, the lens is first optimized to yield the best possible optical quality, then post-processing algorithms are designed to correct residual aberrations. However, it has been shown that optimizing jointly the lens and the post-processing algorithm can yield imaging

systems with better global performance, enhanced capabilities and/or lower complexity [1]–[4].

This joint optimization approach is particularly relevant for designing hybrid imaging systems based on wavefront coding. This technique consists in placing a phase mask in the aperture stop of the lens and applying deconvolution to the obtained image in order to recover contrast. Dowski and

Cathey introduced wavefront coding to increase depth of field by designing phase masks that make the Modulation Transfer Function (MTF) of the lens insensitive to defocus [5]. Different types of masks have then been investigated and their MTF through focus have been optimized in order to make deconvolution easier [6]–[9]. For that purpose, different optimization criteria have been proposed, such as reducing the intensity variations in the focal line [10], securing the absence of zeros in the MTF [8], or implementing a joint optimization approach that consists in maximizing the quality of the restored image by jointly optimizing the phase mask and the deconvolution algorithm [3]. The advantages of wavefront coding have been experimentally demonstrated to simplify the design and complexity of the optical system [11, 12], to reduce focus variations due to temperature changes in imaging systems such as mirror-based telescopes [13], to improve the depth of field of thermal cameras [14] or of microscopes [15, 16], and to increase the acquisition volume for biometric iris imaging system [17].

In this article we use the joint optimization approach to extend, for the first time to our knowledge, the depth of field of a standard frame rate High Definition (HD) $f/1.2$ camera operating in visible and near-infrared spectral range. The design step consists in determining the parameters of a binary phase mask and of a deconvolution algorithm that jointly maximize the restored image quality [3]. The optimal mask is manufactured and inserted in an end-to-end imaging chain including the sensor and a real time deconvolution algorithm implemented on a Field-Programmable Gate Array (FPGA) board. By experimentally estimating the MTF and image quality of the as-built wavefront-coded camera and comparing it to a conventional one, we finally demonstrate an increase in depth of field of a factor 2.5 in operational conditions.

2 DESIGN AND OPTIMIZATION

Let us consider a hybrid imaging system based on a phase mask in the stop plane of the lens. The phase function of the mask depends on a set of parameters that will be denoted φ . In this paper we will consider an annular binary phase mask, composed of six concentric rings whose phase is alternatively zero and π radians (see Figure 1(a)) at a reference wavelength denoted λ_o . The mask is thus defined by the $N - 1$ outer radii, that is $\varphi = \{r_n; n \in \llbracket 1, N - 1 \rrbracket\}$. In such a system the image acquired by the sensor can be modelled as:

$$I_{\psi}^{\varphi}(\mathbf{r}) = h_{\psi}^{\varphi}(\mathbf{r}) * O(\mathbf{r}) + n(\mathbf{r}) \quad (1)$$

where $*$ denotes convolution, \mathbf{r} denotes the image spatial coordinates, $O(\mathbf{r})$ is the perfect image of the scene, $n(\mathbf{r})$ is the acquisition noise and $h_{\psi}^{\varphi}(\mathbf{r})$ denotes the PSF of the lens, that depends on the phase mask parameters φ and on the defocus value ψ defined by:

$$\psi = \frac{\pi R^2}{\lambda_o} \left(\frac{1}{d_o} + \frac{1}{d_i} - \frac{1}{f} \right) \quad (2)$$

where R is the aperture stop radius, λ_o is the reference wavelength, f is the focal length, d_o is the object distance and d_i is the image distance. We take into account the low-pass effect of

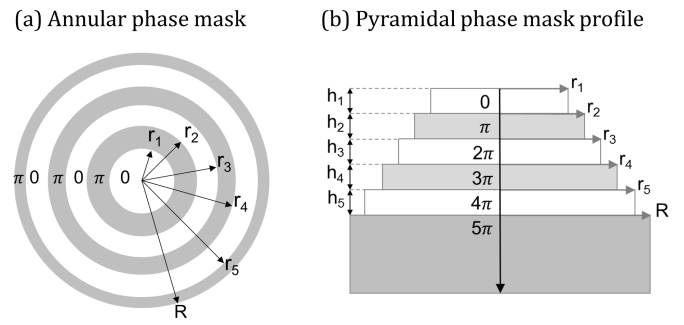


FIG. 1 (a) Six-ring annular phase mask and (b) pyramid shaped mask profile.

pixel integration, but we do not consider aliasing effects. The acquired image is then restored by applying a linear deconvolution filter of impulse response $d^{\varphi}(\mathbf{r})$ as follows:

$$\hat{O}_{\psi}^{\varphi}(\mathbf{r}) = d^{\varphi}(\mathbf{r}) * I_{\psi}^{\varphi}(\mathbf{r}) \quad (3)$$

The goal is to determine the phase mask set of parameters φ_{opt} that minimizes the mean square error (MSE) between the scene and the restored image over a set of K defocus values uniformly distributed in the defocus range $[0; \psi_{max}]$. Following [3], the expression of the MSE for a given value of the defocus ψ is:

$$\begin{aligned} MSE_{\psi}^{\varphi} &= \langle \int |\hat{O}_{\psi}^{\varphi}(\mathbf{r}) - O(\mathbf{r})|^2 d\mathbf{r} \rangle \\ &= \int |\tilde{d}^{\varphi}(\mathbf{v}) \tilde{h}_{\psi}^{\varphi}(\mathbf{v}) - 1|^2 S_{oo}(\mathbf{v}) d\mathbf{v} \\ &\quad + \int |\tilde{d}^{\varphi}(\mathbf{v})|^2 S_{nn}(\mathbf{v}) d\mathbf{v} \end{aligned} \quad (4)$$

where $\langle \rangle$ denotes statistical averaging, the superscript \sim refers to Fourier transform, \mathbf{v} denotes the spatial frequency, $S_{oo}(\mathbf{v})$ denotes the Power Spectral Density (PSD) of the scene O and $S_{nn}(\mathbf{v})$ denotes the PSD of the noise n . The deconvolution filter that minimizes the average MSE over all defocus values, $MSE(\varphi) = \frac{1}{K} \sum_k MSE_{\psi_k}^{\varphi}$, is the following Wiener-like filter [3]:

$$\tilde{d}^{\varphi}(\mathbf{v}) = \frac{\frac{1}{K} \sum_{k=1}^K \tilde{h}_{\psi_k}^{\varphi*}(\mathbf{v})}{\frac{1}{K} \sum_{k=1}^K |\tilde{h}_{\psi_k}^{\varphi}(\mathbf{v})|^2 + \frac{S_{nn}(\mathbf{v})}{S_{oo}(\mathbf{v})}} \quad (5)$$

In practice, we compute this filter using a generic PSD model for $S_{oo}(\mathbf{v})$ [3, 6]. $S_{nn}(\mathbf{v})$ is a constant since the noise n is assumed to be white. We then determine numerically the mask parameters that minimize the maximal MSE over the defocus values, that is,

$$\varphi_{opt} = arg \min_{\varphi} \left\{ \max_k \left[MSE_{\psi_k}^{\varphi} \right] \right\} \quad (6)$$

where the Wiener-like filter in Eq. (5) is used to compute the MSE in Eq. (4).

In this paper, we consider a 1296×972 HD camera with f-number equal to 1.2, a focal length of 20 mm and a pixel pitch of $4.4 \mu\text{m}$. The whole field of view is $16^{\circ} \times 12^{\circ}$. The sensor is sensitive to visible and near infrared radiation. The conventional camera has been designed to resolve details up to 60 lp/mm and provides sharp images for objects from 12 m to infinity. The design of the hybrid imaging system is based on two steps: a traditional optical design of the lens followed

by the joint design of the optical mask and the digital filter following Eq. (6). A model of the lens of this camera and of the phase mask are input in the CodeV® optical design software to perform accurate computation of the optical system PSFs across the object distance range. The optimization loop is managed by the Matlab® computing software.

Our goal is to increase the depth of field by a factor 2.5 so that the system can provide sharp images on a defocus range from 4.8 m to infinity. The desired defocus range is sampled by $K = 3$ defocus values corresponding to 10 km, 9.6 m and 4.8 m. The input sensor level signal-to-noise ratio, defined as:

$$SNR_{in} = 10 \log_{10} \left[\frac{\int S_{oo}(\mathbf{v}) d\mathbf{v}}{\int S_{nn}(\mathbf{v}) d\mathbf{v}} \right] \quad (7)$$

is set to 34 dB. Local optimization using the Nelder-Mead simplex algorithm has been performed on-axis at the reference wavelength $\lambda_0 = 750$ nm. The obtained optimal radii values r_i^{nom} of the phase mask are given in Table 1. We have checked that further increasing the number of rings does not significantly improve the MSE.

i	r_i^{nom} (mm)	r_i^{meas} (mm)	h_i^{nom} (nm)	h_i^{meas} (nm)
1	3.22	3.21 ± 0.01	284	296 ± 10
2	5.25	5.24 ± 0.01	284	289 ± 10
3	5.85	5.83 ± 0.01	284	283 ± 10
4	6.40	6.38 ± 0.01	284	289 ± 10
5	6.94	6.92 ± 0.01	284	303 ± 10

TABLE 1 Nominal radii r_i^{nom} and step heights h_i^{nom} (resp. measured radii r_i^{meas} and heights h_i^{meas}) of each ith concentric step of the pyramidal phase mask.

3 FULL HYBRID IMAGING CHAIN INTEGRATION AND IMPLEMENTATION

The optimal phase mask has been manufactured on a ZnS plate by diamond turning. Since the refractive index of ZnS is 2.32, the mask transition height must be equal to $h = 284$ nm in order to yield transitions of π radians at λ_0 . For ease of realization with diamond turning process, we have manufactured a pyramidal mask whose each step height h_i^{nom} must be equal to h at each transition: it thus consists of a set of six concentric plateaus that lead to phase changes of $0, \pi$ radians, 2π radians, etc. (see Figure 1(b)). The diamond turning technique allows a plate roughness of 10 nm. Thanks to this technique, rings are auto-centred from one to another, and the position of the overall rings depends mainly on the plate position error of ± 0.05 mm. Moreover the manufacturing precision is of ± 15 nm on radii and of ± 15 nm over transition heights. The manufactured phase mask has been characterized using an optical profilometer. The transition heights h_i^{meas} and ring radii r_i^{meas} were estimated by averaging 8 phase mask acquired profiles. The measurement uncertainty is of 10 nm over the transition height and of 0.01 mm on the ring size. Measurement precisions are similar to the manufactured phase plate roughness. Therefore we can consider that standard deviations on radii and height measurements are comparable with measurement precisions values. Results are presented in Table 1. Considering the manufacturing tolerances

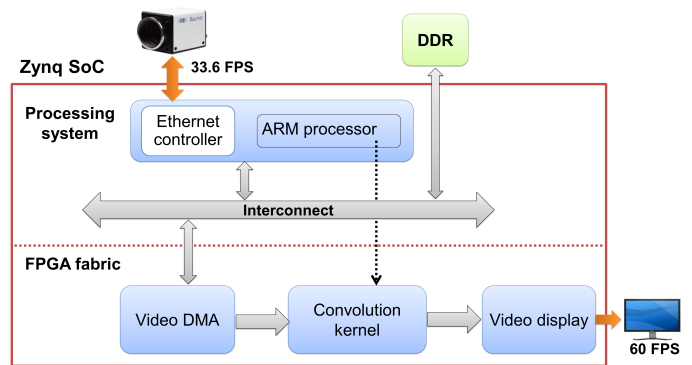


FIG. 2 Architecture mapping for the implementation of real-time HD hybrid imaging chain.

given by diamond turning process, we can say that the manufactured mask complies with specifications. Moreover we have checked that such a pyramidal shape does not introduce significant changes of the mask behavior along the spectral band relatively to the nominal wavelength. The manufactured phase plate is inserted inside the camera lens so that its etched side is located in the stop plane.

The implementation of real-time HD post-processing is depicted in Figure 2. The camera sensor provides 1296×972 pixel images at 33.6 frames per second (FPS). The video stream is sent to a Xilinx ZYNQ ZC702 board, containing a Zynq 7020 System-on-Chip (SoC), which aim is to acquire the video stream, implement the convolution and send the processed video stream to a HD display. The SoC embeds a dual core ARM Cortex A9 @ 866 MHz coupled with an Artix-7 FPGA fabric that provides better performance and energy efficiency than the processor at the expense of programmability. The ARM processor acquires the video stream via an Ethernet link using the GigE Vision interface for high-performance industrial cameras and stores it in DDR memory, while the FPGA is used for retrieving the frame in DDR memory and for the computationally intensive convolution. We use a VHDL Intellectual Property block developed at Thales that convolves the input image with a deconvolution kernel which parameters can be tuned at runtime. The kernel was obtained by computing the inverse Fourier transform of the Wiener-like filter expressed in Eq. (5), and by cropping the result to an 11×11 pixel window. This setup performs the convolution of a 1296×972 pixel image by the 11×11 kernel at up to 119 FPS and therefore allows one to perform real time imaging at the 33.6 FPS camera video rate. As a comparison, the same computation performed by a general purpose processor would barely reach 1 FPS. The total FPGA power consumption has been measured to be 600 mW including 430 mW for the convolution itself. It has to be compared with the nominal 5 W power consumption of the camera. The deconvolved image is finally displayed on a HD screen at 60 FPS.

The whole imaging chain is represented in Figure 3. Two identical cameras are set side by side. One of them includes the phase mask in the stop of its lens and is connected to the FPGA board for deconvolution. It will be called in the following the 'hybrid camera'. The other camera includes in its stop plane a simple ZnS plate with the same average thickness as the phase

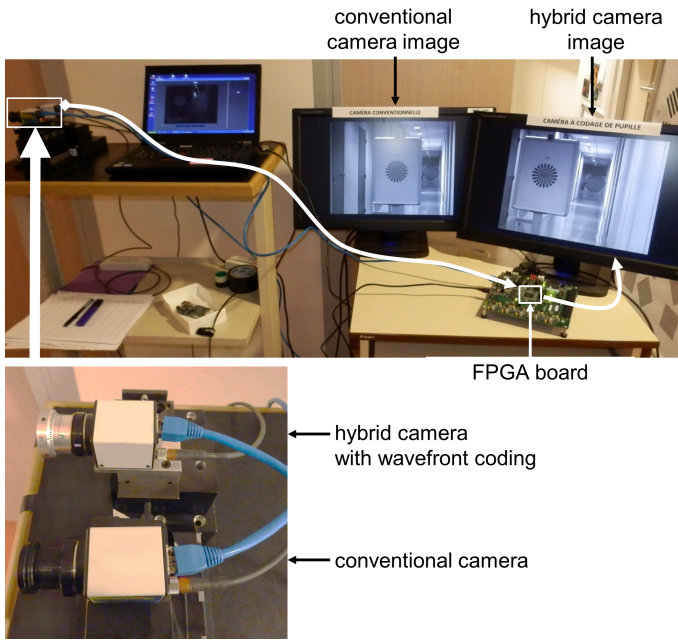


FIG. 3 Experimental set-up for the real-time image comparison between the conventional and the hybrid cameras.

mask. It will be called in the following the ‘conventional camera’. Thanks to this setup, we can compare images of the same scene acquired at the same time by both cameras.

4 EXPERIMENTAL CHARACTERIZATION AND IMAGING DEMONSTRATION

To characterize and compare imaging performance of both cameras, we first estimated their polychromatic MTF. We conducted measurement indoor at the three following object distances: 4.8 m, 9.6 m and 47 m which was the maximum reachable distance. The MTF measurement method consists in estimating the horizontal MTF from vertical bar targets corresponding to spatial frequencies on the sensor of 10 lp/mm, 20 lp/mm, 30 lp/mm, 40 lp/mm, 50 lp/mm and 60 lp/mm. All these targets have similar spatial extensions in order to minimize phasing error on MTF estimation. For calibration needs, we acquired at each distance a full white target image and a full black target image. The target image has been normalized with respect to the white and the black target images as follows:

$$t_0(\mathbf{r}) = \frac{t(\mathbf{r}) - b(\mathbf{r})}{w(\mathbf{r}) - b(\mathbf{r})} \quad (8)$$

where $t(\mathbf{r})$ denotes the bar target image at a given frequency, $w(\mathbf{r})$ denotes the white image, $b(\mathbf{r})$ denotes the black image, and $t_0(\mathbf{r})$ denotes the calibrated bar target image. Horizontal profiles were extracted from the calibrated image and averaged to produce one single profile with reduced noise. To avoid side effects caused by its finite extension, the averaged profile has been apodized using a Hamming window function. The apodized profile is denoted $\{p_k\}_{k=1..P}$ where P is the number of profile samples. The MTF was then estimated in the following way, when $\nu_0 \neq 0$:

$$MTF(\nu_0) = \frac{\pi}{2} \cdot \frac{\sum_{k=0}^P p_k e^{-i2\pi p_k \nu_0}}{\sum_{k=0}^P p_k} \quad (9)$$

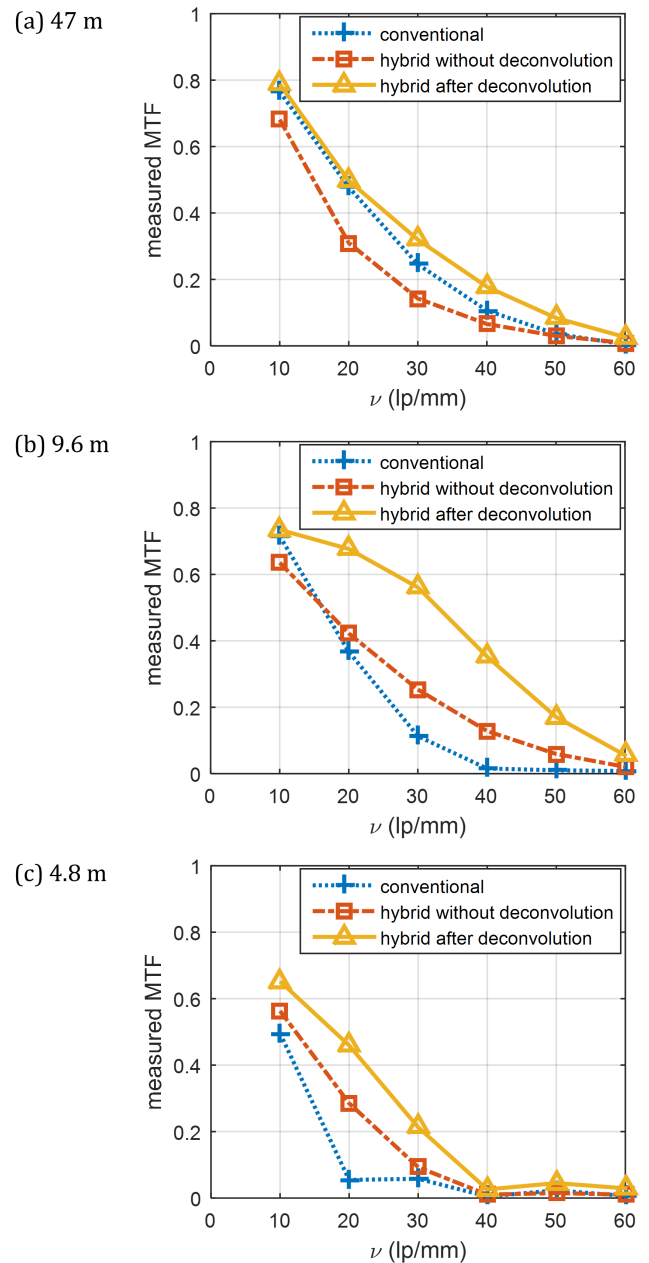


FIG. 4 Measured MTF of the conventional and the hybrid camera from bar targets positioned at (a) 47 m, (b) 9.6 m and (c) 4.8 m.

For conventional and hybrid cameras, the focus have been set respectively at 24 m and 9.6 m in order to provide the most optimized condition for depth of focus and image quality. All experiments have been carried out at full aperture ($f/1.2$). Measured MTF for both cameras at the three object distances are represented in Figure 4. As expected, the conventional camera MTF (blue curves) dramatically decrease as the object distance gets smaller. The MTF of the hybrid camera without deconvolution (red curves) is lower than the conventional one at 47 m, but as expected, it varies much less with the focus distance, which will allow us to perform deconvolution with a single filter. Using the same bar target-based method, we measured the MTF of the hybrid camera with deconvolution. The results are showed in Figure 4 (yellow curves). It can be noticed that the MTF of the hybrid camera is above that of the conventional camera at each frequency for all object positions, especially for the closest ones. As the deconvolution

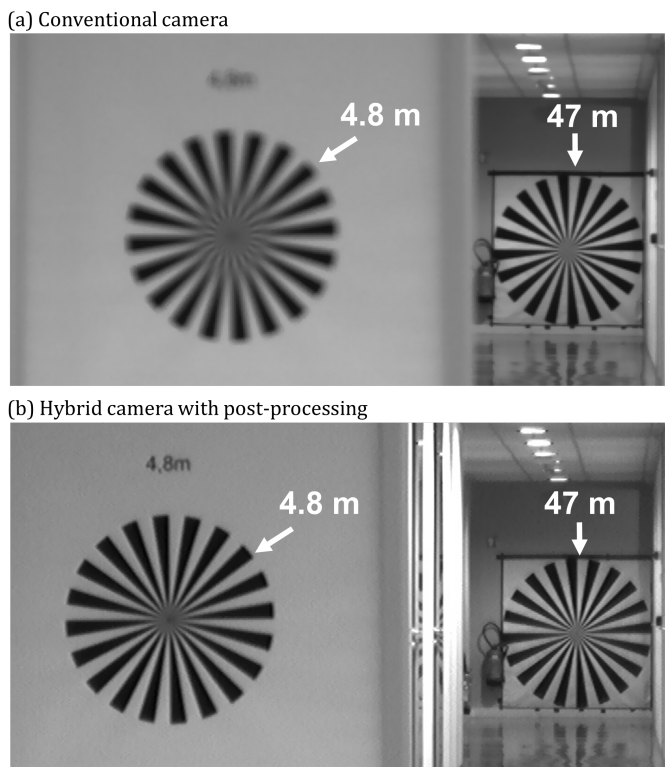


FIG. 5 Spoke targets positioned at 4.8 m and 47 m imaged by (a) the conventional camera focused on the farthest target at 47 m, and (b) the hybrid camera.

filter was designed to restore frequencies that have been attenuated by the lens with phase mask, the ideal hybrid system without noise would have a MTF equal to one over the frequency range and would be therefore higher than the MTF of a diffraction-limited lens without post-processing. In practical conditions, the post-processed hybrid camera MTF does not reach 1 but has significantly higher values especially at 9.6 m than the conventional system as shown in Figure 4. This validates the performance of the hybrid camera.

As a further illustration, we compare in Figure 5 images of the same scene acquired by both cameras at full aperture. The scene is composed of two spoke targets positioned at 4.8 m and at 47 m. These targets contain all spatial frequencies, from 5 lp/mm at the periphery to infinity at the centre, and therefore allow one to observe continuous contrast variations with respect to spatial frequency. It can be noticed that the hybrid system allows to overcome the conventional camera incapacity to provide sharp images simultaneously at 4.8 m and 47 m for a given focus setting. For the conventional camera, the nearest spoke target is significantly blurred and has two contrast inversions indicating that the MTF goes to zero at least twice as it is confirmed by the measured MTF in Figure 4(c) (blue curve). For the hybrid system instead, both targets are sharp and have similar image quality.

5 CONCLUSION

In conclusion, we experimentally demonstrated the increase in depth of field of a $f/1.2$ HD visible and near infrared camera using wavefront coding. We designed and built an end-to-end hybrid imaging system combining an optimized six ring

pyramidal phase mask placed in the aperture stop of the lens and a digital deconvolution algorithm. The performance of this hybrid camera has been compared to that of a conventional camera based on the same conventional optical design by performing MTF measurements and imaging experiments. The results show an increase in depth of field of a factor 2.5. This is, to the best of our knowledge, the first time that such an increase of performance using wavefront coding is designed, implemented and characterized for a system providing real time High Definition images with a very small f -number.

This result shows that it is possible to easily increase the depth of field of traditionally designed optical systems by simply inserting a mask in the stop plane and using FPGA-based deconvolution filter while keeping the optical system as compact and lightweight as the traditional one. This could lead to many applications in such domains as embedded imaging systems, handheld or head-mounted devices, where compactness and weight are key parameters. For example, an attractive application of this approach would be to increase the performance of security and surveillance equipment at low light level while reducing their cost and weight by avoiding mechanical parts for focusing.

References

- [1] W. Cathey, and E. Dowski, "New paradigm for imaging systems," *Appl. Optics* **41**, 6080-6092 (2002).
- [2] D. Stork, and M. Robinson, "Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems," *Appl. Optics* **47**, B64-B75 (2008).
- [3] F. Diaz, F. Goudail, B. Loiseaux, and J. Huignard, "Increase in depth of field taking into account deconvolution by optimization of pupil mask," *Opt. Lett.* **34**, 2970-2972 (2009).
- [4] T. Vettenburg, and A. Harvey, "Holistic optical-digital hybrid-imaging design: wide-field reflective imaging," *Appl. Optics* **52**, 3931-3936 (2013).
- [5] E. Dowski, and W. Cathey, "Extended depth of field through wavefront coding," *Appl. Optics* **34**, 1859-1866 (1995).
- [6] F. Diaz, F. Goudail, B. Loiseaux, and J. Huignard, "Comparison of depth-of-focus-enhancing pupil masks based on a signal-to-noise-ratio criterion after deconvolution," *J. Opt. Soc. Am. A* **27**, 2123-2131 (2010).
- [7] O. Palillero-Sandoval, J. Félix Aguilar, and L. Berriel-Valdos, "Phase mask coded with the superposition of four Zernike polynomials for extending the depth of field in an imaging system," *Appl. Optics* **53**, 4033-4038 (2014).
- [8] R. N. Zahreddine, and C. J. Cogswell, "Binary phase modulated partitioned pupils for extended depth of field," in *Proceedings to Imaging and Applied Optics 2015*, IT4A.2 (Optical Society of America, Arlington, 2015).
- [9] V. Le, Z. Fan, and Q. Duong, "To extend the depth of field by using the asymmetrical phase mask and its conjugation phase mask in wavefront coding imaging systems," *Appl. Optics* **54**, 3630-3634 (2015).
- [10] L. Liu, F. Diaz, L. Wang, B. Loiseaux, J. Huignard, C. Sheppard, and N. Chen, "Superresolution along extended depth of focus with binary-phase filters for the Gaussian beam," *J. Opt. Soc. Am. A* **25**, 2095-2101 (2008).

- [11] M. Demenikov, E. Findlay, and A. Harvey, "Miniaturization of zoom lenses with a single moving element," *Opt. Express* **17**, 6118–6127 (2009).
- [12] G. Muyo, A. Singh, M. Andersson, D. Huckridge, A. Wood, and A. Harvey, "Infrared imaging with a wavefront-coded singlet lens," *Opt. Express* **17**, 21118–21123 (2009).
- [13] X. Guo, L. Dong, Y. Zhao, W. Jia, L. Kong, Y. Wu, and B. Li, "Imaging and image restoration of an on-axis three-mirror Cassegrain system with wavefront coding technology," *Appl. Optics* **54**, 2798–2805 (2015).
- [14] F. Diaz, M. Lee, X. Rejeaunier, G. Lehoucq, F. Goudail, B. Loiseaux, S. Bansropun, et al., "Real-time increase in depth of field of an uncooled thermal camera using several phase-mask technologies," *Opt. Lett.* **36**, 418–420 (2011).
- [15] T. Zhao, T. Mauger, and G. Li, "Optimization of wavefront-coded infinity-corrected microscope systems with extended depth of field," *Biomed. Opt. Express* **4**, 1464–1471 (2013).
- [16] R. Zahreddine, and C. Cogswell, "Total variation regularized deconvolution for extended depth of field microscopy," *Appl. Optics* **54**, 2244–2254 (2015).
- [17] Y. Q. He, J. Q. Li, J. Pan, and Y. J. Li, "Optimization of phase mask based iris imaging system through the optical characteristics," *Proc. SPIE* **8711**, 871107 (2013).