Computational All-in-Focus Imaging using an Optical Phase Mask

Harel Haim*, Alex Bronstein and Emanuel Marom

Iby and Eladar Fleischmann Faculty of Engineering,
Tel Aviv University, Tel Aviv 69978, Israel
Harelhai@mail.tau.ac.il

Abstract: A method for extended depth of field imaging based on image acquisition through a thin binary phase plate followed by fast automatic computational post-processing is presented. By placing a wavelength dependent optical mask inside the pupil of a conventional camera lens, one acquires a unique response for each of the three main color channels, which adds valuable information that allows blind reconstruction of blurred images without the need of an iterative search process for estimating the blurring kernel. The presented simulation as well as capture of a real life scene show how acquiring a one-shot image focused at a single plane, enable generating a de-blurred scene over an extended range in space.

©2015 Optical Society of America

OCIS codes: (110.1758) Computational imaging; (110.1455) Blind deconvolution; (110.3010) Image reconstruction techniques; (100.3020) Image reconstruction-restoration; (100.3190) Inverse problems

References and links

1. B. Milgrom, N. Konforti, M. A. Golub, and E. Marom, "Pupil coding masks for imaging polychromatic scenes with high resolution and extended depth of field," *Opt Express*, **18** (15), 15569–15584, (2010).
2. B. Milgrom, N. Konforti, M. A. Golub, and E. Marom, "Novel approach for extending the depth of field of Barcode decoders by using RGB channels of information," *Opt Express*, **18** (16), 17027–17039, (2010).
3. J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in Astronomy: A Review," *Publ. Astron. Soc. Pacific*, The University of Chicago Press, **114** (800), 1051–1069, (2002).
4. M. Elad, *Sparse and redundant representations : from theory to applications in signal and image processing*. (Springer, 2010).
5. M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Trans. Image Process.*, **15** (12), 3736–3745, (2006).
6. M. J. Fadili, J. L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *Comput. J.*, **52** (1), 64–79, (2009).
7. F. Couzinie-Devy, J. Mairal, F. Bach, and J. Ponce, "Dictionary learning for deblurring and digital zoom," *arXiv Prepr. arXiv1110.0957*, (2011).
8. M. S. C. Almeida and L. B. Almeida, "Blind and semi-blind deblurring of natural images.," *IEEE Trans. Image Process.*, **19** (1), 36–52, (2010).
9. Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *ACM Transactions on Graphics (TOG)*, ACM, **27** (3), pp. 73.
10. Z. Hu, J. Bin Huang, and M. H. Yang, "Single image deblurring with adaptive dictionary learning," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, IEEE, pp. 1169–1172.
11. D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE, pp. 233–240.
12. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep. CSTR*, **2** (11), (2005).
13. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, **26** (3), 70, (2007).
14. F. Guichard, H.-P. Nguyen, R. Tessières, M. Pyanet, I. Tarchouna, and F. Cao, "Extended depth-of-field using sharpness transport across color channels," in *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics (2009), pp. 72500N–72500N–12.
15. J. W. Goodman, *Introduction to Fourier optics*, 2nd ed. (McGraw-Hill, 1996).

16. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, **20** (1), 33–61, (1998).

1. Introduction

The increasing popularity of social networks combined with the affordability of high-quality CMOS sensors have transformed digital cameras into an integral part of the way we communicate. The demand for quality high-resolution cameras, specifically for smart phones and hand-held devices, created a competitive market that is constantly looking for the next big thing.

Digital image quality is determined by the imaging system properties and the focal plane array sensor. With the increase in pixel number and density, imaging system resolution is now mostly bound by optical system limitations. Form factor challenges imposed by the smart phone market makes it very difficult to improve the image quality via standard optical solutions, and therefore most of the advancements in recent years have shifted to the domain of image processing.

One of the most challenging issues in imaging systems is the restoration of out-of-focus (OOF) images. The problem is notoriously ill-posed, since information is lost in the process. As presented in [1], a symmetric binary phase mask offers a low-cost optical solution to this problem, providing acceptable quality for machine vision applications. By using the right phase mask, one can increase the camera's depth of field (DOF) for different uses such as barcode reading and face detection, as well as for other machine vision applications. A novel approach initiated in our group [2], proposed using an RGB phase mask, whereby one gets different responses for the red, green and blue channels, resulting in a simultaneous acquisition of three perfectly registered images, each with a different out-of-focus characteristic. Under the assumption that sharp edges and other high-frequency features characterizing a well-focused image are present in all color channels, the joint analysis of the three channels captured through the phase mask allows to extend the system DOF.

Even though the phase mask solution presented [1] is inexpensive and suits most machine vision algorithms with no additional computational power, it limits image analysis applications, since visually the output color image is poorer than that of an image taken without the mask, in particular, for in-focus condition. In this paper, we present a method for fusing the three RGB channels into a single color image with extended DOF and improved appearance via a specially tailored efficient post-processing algorithm.

There exists a wealth of literature dedicated to purely computational approaches to image deblurring and deconvolution [3]. The dominant models, increasingly popular in recent years, are flavors of sparse and redundant representations [4], which have been proved as a powerful tool for image processing, compression, and analysis. It is now well-established that small patches from a natural image can be represented as a linear combination of only a few atoms in an appropriately constructed over-complete (redundant) dictionary. This constitutes a powerful prior fact that has been successfully employed to regularize numerous otherwise ill-posed image processing and restoration tasks [5–7].

Existing image deblurring techniques assume that the blurring kernel is known [7] or use an expectation maximization-type of approach to blindly estimate the kernel from the data [8–11]. However, by using conventional optics, information about sharp image features is irreversibly lost, posing inherent limitations on any computational techniques. We propose to take advantage of the differences between the responses of the RGB channels, when the imaging system includes a special phase mask, in order to restore the blurred image in a single run, and without prior knowledge of the blurring kernel. Moreover, our system can handle scenes consisting of features located at several depth of field, thus having different blur characteristics.

2. Related work

The marriage between optics and image processing has received significant attention in the past decade, to the extent that it is now singled out as a new field of *computational photography*. Light field cameras [12] present some interesting abilities concerning focus manipulation and improving DOF. Even so, light field cameras suffer from low resolution, bulkiness, and noise and require a unique design, which makes it harder to integrate them into existing system (smartphones, laptop cameras…).

Another related work by Levin et al. [13] used a coded aperture with a conventional camera to capture an all-in-focus image after proper electronic processing. The downside of this approach is the fact that being an amplitude mask, this coded aperture blocks around 50% of the light making the system much more sensitive to noise especially in low-light conditions. A different approach [14] utilizes a lens with deliberately high chromatic aberration to achieve color separation, but requires special lens design for each system. Although similar in principle, our design achieves color separation with a simple thin phase mask, which can be incorporated in any conventional camera with minimal optical losses; the acquisition is followed by a unique processing stage.

3. Out-of-focus imaging and its effects on image quality

Classical lens systems are commonly used for capturing images of "natural" scenes in incoherent light illumination. The properties of the imaging system are usually analyzed in the spatial frequency domain using the Fourier transform representation. The modulation transfer function (MTF) essentially provides the attenuation factor (or contrast) of all spatial frequencies passing through the system.

An imaging system acquiring an Out-of Focus (OOF) object, suffers from aberrations, in particular blur, that degrade the image quality, meaning low contrast , loss of sharpness and even loss of information. In digital systems, image blur (main OOF effect) is not observable as long as the image size of a point source in the object plane is smaller than the pixel size in the detector plane.

The OOF error is analytically a wave-front error, namely a quadratic phase error in the pupil plane [15]. In case of a circular aperture with radius *R,* we define the defocus parameter as:

$$\Psi = \frac{\pi R^2}{\lambda}\left(\frac{1}{z_0}+\frac{1}{z_{img}}-\frac{1}{f}\right)=\frac{\pi R^2}{\lambda}\left(\frac{1}{z_{img}}-\frac{1}{z_i}\right)=\frac{\pi R^2}{\lambda}\left(\frac{1}{z_o}-\frac{1}{z_n}\right) \qquad (1)$$

where z_{img} is the detector plane location when the object is in the nominal position z_n, z_i is the ideal image plane for an object located at z_o, and λ is the optical wavelength.

The defocus parameter Ψ measures the maximum phase error at the aperture edge. For $\Psi > 1$ the image will experience contrast loss and for $\Psi > 4$ it will experience information loss and even reversal of contrast at some frequencies, as readily observed in Fig. 1.

For a circular clear aperture, the diffraction limit maximum spatial frequency (or the cut-off frequency) is given by $f_c = 2R/\lambda z_i$. As the aperture size increases the resolution of the optical system rises while, at the same time, reducing the DOF of the system. The increase in the aperture size also increases the amount of light collected by the system, thus improving its SNR.

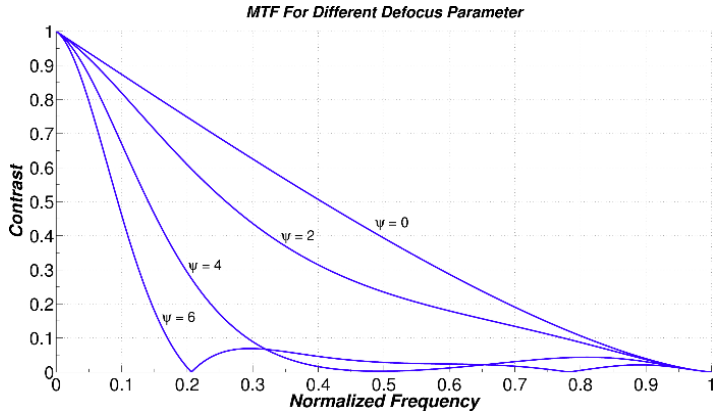

Fig. 1 - MTF for a circular aperture for different value of the defocus parameter.

4. Pupil coding phase mask for extended DOF imaging

Radially symmetric binary optical phase masks have been proposed for overcoming limitations set by OOF imaging [1]. Those masks are composed of one or several rings providing a predetermined phase-shift (usually, exactly π). However, such phase mask provides the exact desired phase shift for a single wavelength only, while the phase shift for other wavelengths changes accordingly. A phase mask, incorporated in the optical train is meant to compensate for the OOF phase error by adding a constant phase shift near the aperture edge. Careful design of the phase level along a radial ring results in an increased DOF. The drawback is that one gets a reduced contrast level when the image is in-focus. For computer vision applications, as for instance for barcode reading, this is not a problem, but for visual observation the resulting image quality is unacceptable.

Milgrom et al [2] proposed the use of a special RGB phase mask that exhibits significantly different response in the three major color channels R, G and B. It has been shown that each channel provides best performance for different depth regions, so that the three channels jointly provide an extended DOF. This mask provided a phase shift of 3π for a blue wavelength and consequently 2π phase shift for a red wavelength. Therefore, this mask did not affect the red channel imaging performance. The challenge addressed in the present paper is the fusion of the RGB channels into an improved color image, obtainable via electronic post-processing.

It is instructive to examine the contrast level of a single spatial frequency (say $f_c / 4$) for different Ψ values. Figure 2 provides a comparison for three cases: clear aperture (Fig. 2a), 3π mask (Fig. 2b) and 4π mask (Fig. 2c). The 4π mask exhibits larger DOF than that of the 3π mask and provides wider separation of the three color channels. This effect is crucial for the post processing stage as it will be described in section 6. It is illustrative to evaluate the distances corresponding to the Ψ values used in Fig. 2. For instance, for an iPhone 6 camera (having a 1.8mm aperture), when focused at an object nominally located 1.4m away, the Ψ factors ranging from (-4) to 8 correspond respectively to object locations extending from infinity up to 50cm from the camera.

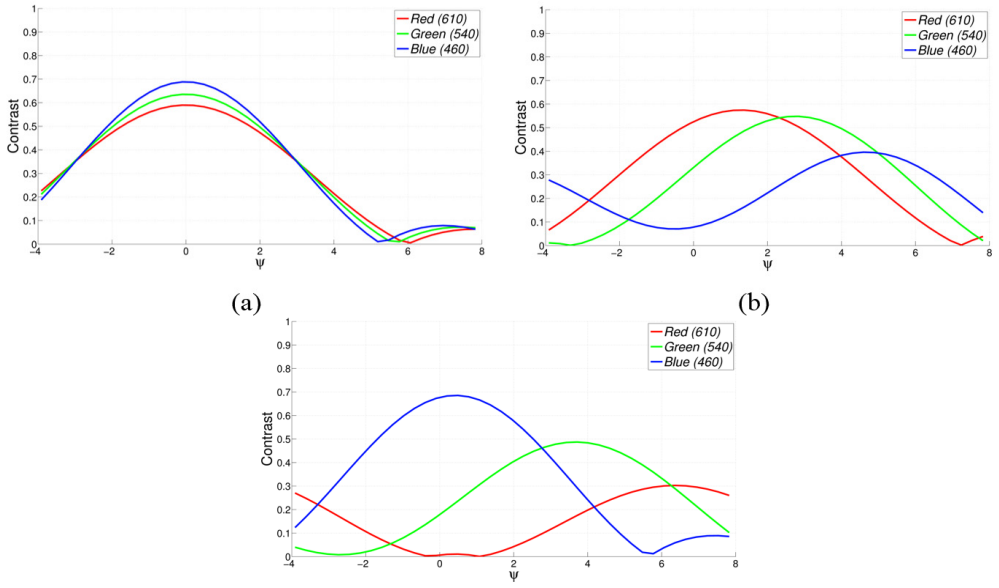

Fig. 2 - Comparison between simulated MTFs of a single spatial frequency ($f_c / 4$) as a function of OOF factor Ψ in case of clear aperture (a) aperture equipped with a phase mask of 3π (b) and 4π (c).

5. Sparse model for non-blind image deblurring using dictionaries pair

Sparse representation has proven to be a strong prior for non-blind image deblurring solutions, were the blurring kernel is known, as well as for blind ones. We base our blind solution on a non-blind one in order to produce a fast post-processing solution.

The signal x is said to admit a sparse representation (or, more accurately, approximation) over $\mathbf{D}$ if one can find a vector $\hat{\alpha} \in \mathbb{R}^N$ with only a few non-zero coefficients, such that $x \approx D\hat{\alpha}$. The sparse representation pursuit problem can be cast as the ℓ_0 pseudo-norm minimization,

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\| x - D\alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_0 \tag{2}$$

where $\left\| \alpha \right\|_0$ counts the number of non-zeros in the vector α, and λ controls the relative importance of the ℓ_0 regularization term over the ℓ_2 data fitting term.

While the sparse representation pursuit (Eq. (2)) is computationally intractable, it can be efficiently approximated by several known techniques such as orthogonal matching pursuit (OMP) [16]. The dictionary D can be constructed axiomatically based on image transforms such as DCT or wavelet, or learned from a training set sampled from representative images [4]. Here, we adopt the latter approach to construct a structured dictionary for the representation of 8×8 patches represented as 64-dimensional vectors.

Assuming we have a sharp image x that has been blurred by a known kernel h the blurred image y can be described as:

$$y = h * x + \eta \tag{3}$$

where η is additive Gaussian noise and '$*$' refers to the convolution operator. In the scope of this work we assumed that $\eta < 1$. Consider a patch from the blurred image represented in

raster order as a column vector $y_i \in \mathbb{R}^m$ and a blurred dictionary $D_b \in \mathbb{R}^{m \times N}$ composed of N m-dimensional atoms generated from the sharp atoms in dictionary D_s, such that $D_b = h * D$.

Using OMP we attempt to solve the following problem:

$$\hat{\alpha}_i = \underset{\alpha}{\operatorname{argmin}} \left\| y_i - D_b \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_0 \tag{4}$$

where $\hat{\alpha}_i \in \mathbb{R}^N$ is the is the vector corresponding to the *i*-th patch y_i. Solution of Eq. (4) produces the sparse code coefficients $\hat{\alpha}$ of the blurred image as a linear combination of atoms from the blurred dictionary D_b. The sharp patches can be recovered by $x_i \approx D_s \hat{\alpha}_i$ .This process implies that for all *i*, $y_i \approx h * D_s \hat{\alpha}_i \approx D_b \hat{\alpha}_i$.

6. Blind color image deblurring using phase mask

The diversity exhibited in the color information when using a phase mask, provides additional information on the captured scene that allows applying the non-blind deblurring technique described in the previous section. In the case of blind image deblurring, the blur kernel is unknown; moreover it varies with the object depth location.

Many studies dealt with this problem, with limited success. For instance, using different iterative processes [8–11], one tries to estimate the blurring kernel so that image restoration can be thereafter achieved. Reconstruction processes usually require high computational complexity, which limits their use for many real-time applications. Our approach based on using an optical phase mask allows restoring the image without needing any iterative process to estimate the blurring kernel. Furthermore, most of blind deblurring algorithms are constructed with the explicit assumption that the input image is localized at a single depth position, such that the blur can be described as a convolution with a spatially constant kernel. In real life scenes, however, the objects are at different distances, and the blur changes abruptly when crossing the object boundaries. Many natural scenes can be approximated by a 2.5D world assumption asserting that a scene comprises a plurality of objects at different depths, yet each object is assumed planar and perpendicular to the optical axis. Under this assumption, the blur can be modeled as the convolution with a piece-wise constant blur kernel defined by the defocus parameter Ψ of that particular object. We also assume that the blurring kernel is affected only by the defocus parameter ($\Psi = 0$ to 8), ignoring other sources of blur, e.g., due to camera or scene motion.

For this application, we propose to construct a number of sub-dictionaries using different blurring kernels for different defocus parameters ($\Psi = 0$ to 8). Assuming that a step of $\Delta\Psi = 1$ is discriminative enough, one needs to construct nine sub-dictionaries and then concatenate them into a single "multi-focus dictionary" D_{MDF}:

$$D_{MFD} = \left[D_s * h_0; D_s * h_1; ...; D_s * h_8 \right] = \left[D_{b1}; ...; D_{b8} \right] \tag{5}$$

Color images are decomposed into 8×8×3 patches represented as 192-dimensional vectors containing the RGB information for each patch. The blurred MFD D_{MDF} is composed of the sub-dictionaries generated from D_s using different blurring kernels. One can now describe the blurry image using D_{MDF}, but without knowing the blurring kernel; note, that in general there is no assurance that elements from the correct sub-dictionary will be selected in the pursuit process.

The OMP process chooses elements from the dictionary that best matches the input patch, based on largest inner product. Using the RGB phase mask, the blurred dictionary varies strongly for different kernels since the response of the imaging system is very different for each color An input patch (or vector) from an imaging system with the RGB mask will also exhibit different response for each color. The response will be unique for each kernel and

therefore the input vector will most likely associate with vectors from D_{bi} that experience the same blurring process.

Comparing our algorithm with an open access state of art algorithm, provided by Krishnan et al. [11], our process produced better results when applied to natural images out of the Kodak dataset[1]. We also run the process on texture images (Colored Brodatz Texture database[2]) and observed similar performance, meaning that the MFD dictionary will work on almost any natural scene. The example in Fig. 3 shows how an OOF image taken with a conventional clear aperture (Fig. 3b) cannot be restored using Krishnan algorithm [11] (Fig. 3c). When using a phase mask the image is visually better (Fig. 3d) than that of a clear pupil, but the restoration process applying Krishnan [11] still introduces strong artifacts (Fig. 3e). Applying our process on the image taken with the phase mask one gets an improved sharp image (Fig. 3f).

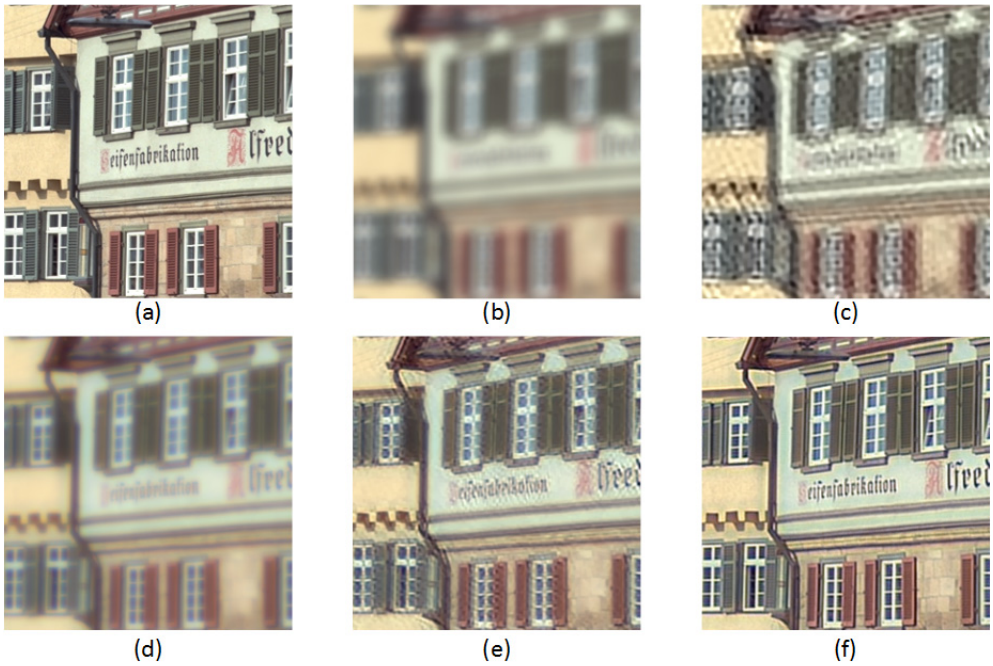

Fig. 3 – Example of simulated image blurring and restoration. (a) Original image from the KODAK dataset ; (b) out of focus with clear aperture; (c) Deblurring of (b) using Krishnan [11]; (d) out of focus with phase mask (e) Deblurring of (d) using Krishnan [11]; (f) – Deblurring of (d) using our new process.

7. Multi-focus scene deblurring

As indicated in the previous section, for natural depth scenes one cannot assume that the image is blurred by a single blurring kernel. Our process analyzes small patches and not the whole image; therefore the restoration process is applied to every region inside the image frame independent on the restoration process of other regions in the frame. This feature is what allows implementing our process directly in one step on a 2.5D scene.

To demonstrate the process let's assume a 2.5D scene with four objects each located at a different distance from the camera as shown in Fig. 4. The top image shows the simulation of capturing a scene using a conventional camera focused on the background. As expected, other

[1] http://r0k.us/graphics/kodak/
[2] http://multibandtexture.recherche.usherbrooke.ca/colored%20_brodatz.html

objects in the scene are blurred according to their distance from the focus point. The bottom image presents the result obtained with an imaging system comprising a phase mask followed by our post process blind restoration scheme. One can notice that all objects were restored without noticeable artifacts.

To reduce running time we examined the minimal number of sub-dictionaries that will still provide good results in comparison to the results shown in Fig. 4. To cover the full range of OOF factors inside the scene one will need to use at list three sub-dictionaries (corresponding to $\Psi = 0, 4, 8$) without losing any noticeable quality.

Fig. 4 - Simulated 2.5D scene with four objects each located at a different distance from the camera coresponding to $\Psi = 0$ (background buildings) to $\Psi = 8$ (the woman on the right) – Conventional imaging were focused on the background (top). Imaging with our system using a phase mask and blind post-processing (bottom).

8. Experimental results

A proof-of-concept experimental system aiming to test the approach described in this paper was then carried out. The system consisted of a commercial CCD camera and a 16mm commercial lens onto which we were able to insert a phase mask. The mask was manufactured on a Soda-Lime substrate onto which a phase pattern was etched in the center to provide the necessary phase shift. A view of the scene setup, onto which we marked the Ψ value corresponding to the object position is shown in Fig. 5.

A comparison between a conventional camera and our method is presented in Fig. 6. Both images were captured in the exact same lighting conditions and exposure time. The left image in Fig. 6 shows the captured scene with a conventional lens (clear aperture) that was focused on the background poster. The right image in Fig. 6 shows the results of capturing the same

scene using an aperture with a phase mask (see Fig. 5) followed by our post processing algorithm. One can notice that using the proposed method one restores an image with all objects in focus. The zoomed sections in Fig. 6 allow the viewer to observe the restored image quantitatively with the use of the resolution target object. Notice that the conventional captured Rosetta image exhibits low contrast or even contrast reversal, as opposed to the sharp restoration results one gets using our system.

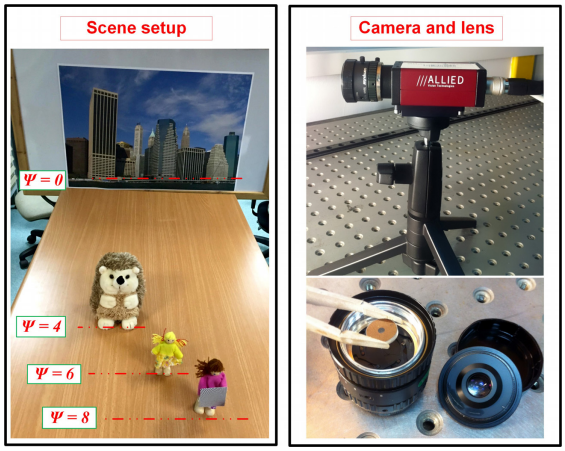

Fig. 5 – Experimental set-up: Left - Scene line-up including the relative OOF factor for each plane. Right - View of camera and mask insertion in the lens assembly.

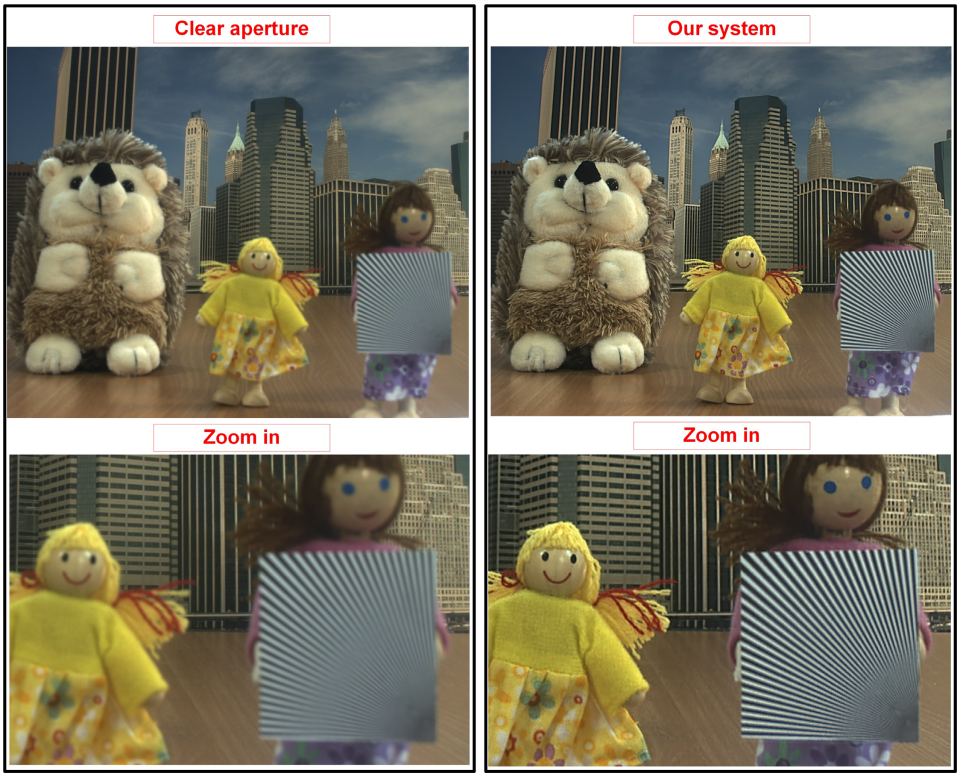

Fig. 6 - Experimental results - Imaging with a conventional clear aperture (left) and imaging with a phase mask and our post processing restoration (right).

9. Conclusions

In this paper, we have presented the foundation for extended depth-of field system, based on a modified conventional optical imaging system equipped with a special thin binary phase mask, followed by an electronic post-processing stage. The processing is based on simple dictionary based image deblurring algorithm. It can be implemented in real-time systems using field programmable gate arrays for lab demonstration or dedicated hardware (ASIC, etc.) in future implementations. Our method was tested successfully on real life natural depth scenes with no need of prior knowledge about the scene composition or user intervention. The experimental results provide added validity of our method. Further work will include depth estimation of various objects that compose the scene which will provide additional features, such as re-focusing capabilities (changing the focus point and DOF of the captured image).