

Perceptually-Optimized Coded Apertures for Defocus Deblurring

Belen Masia, Lara Presa, Adrian Corrales and Diego Gutierrez

Universidad de Zaragoza, Spain

Abstract

The field of computational photography, and in particular the design and implementation of coded apertures, has yielded impressive results in the last years. In this paper we introduce perceptually-optimized coded apertures for defocused deblurring. We obtain near-optimal apertures by means of optimization, with a novel evaluation function that includes two existing image quality perceptual metrics. These metrics favor results where errors in the final deblurred images will not be perceived by a human observer. Our work improves the results obtained with a similar approach that only takes into account the L2 metric in the evaluation function.

Categories and Subject Descriptors (according to ACM CCS): I.4.3 [Image Processing and Computer Vision]: Enhancement—Sharpening and deblurring

1. Introduction

In the past few years, the field of computational photography has yielded spectacular advances in the imaging process. One strategy is to code the light information in novel ways before it reaches the sensor, in order to decode it later and obtain an enhanced or extended representation of the scene being captured. This can be accomplished for instance by using structured lighting, new optical devices or modulated apertures or shutters. In this work we focus on *coded apertures*. These are masks obtained by means of computational algorithms which, placed at the camera lens, encode the defocus blur in order to better preserve high frequencies in the original image. They can be seen as an array of multiple ideal pinhole apertures (with infinite depth and no chromatic aberration), whose location on the 2D mask is determined computationally. Decoding the overlap of all pinhole images yields the final image.

Some existing works interpret the resulting coded blur attempting to recover *depth from defocus*. Given the nature of the blur as explained by simple geometrical optics, this approach imposes a multi-layered representation of the scene being depicted. While there is plenty of interesting on-going research in that direction, in this paper we limit ourselves to the problem of *defocus deblurring*: we aim to obtain good coded apertures that allow us to recover a sharp image from its blurred original version. We follow standard approaches

and pose the imaging process as a convolution between the original scene being captured and the blur kernel (plus a noise function). In principle, this would lead to a blind deconvolution problem, given that the such blur kernel is usually not known. Assuming no motion blur nor camera shake, this kernel is reduced to the point spread function of the optical system. Traditional circular apertures, however, have a very poor response in the frequency domain: not only do they lose energy at high frequencies, but they exhibit multiple zero-crossings as well; it is thus impossible to recover information at such frequencies during deconvolution.

Inspired by previous works [ZN09], we rely on the average power spectra of natural images to guide an optimization problem, solved by means of genetic algorithms. Our main contribution is the use of two existing image quality perceptual metrics during the computation of the apertures; this leads to a new evaluation function that minimizes errors in the deconvolved images that are predicted to be perceived by a human observer. Our results show better performance compared to similar approaches that only make use of the L2 metric in the evaluation function. Additionally, we explore the possibility of computing non-binary masks, and find a trade-off between ringing artifacts and sharpness in the deconvolved images. Our work demonstrates a novel example of applying perceptual metrics in different contexts; as these perceptual metrics evolve and become more sophisticated,

some existing algorithms may be revisited and benefit from them.

2. Previous Work

Coded apertures have been traditionally used in astronomy, coding the direction of incoming rays as an alternative to focusing imaging techniques which rely on lenses [ItZ92]. Possibly the most popular patterns were the MURA patterns (Modified Uniformly Redundant Array) [GF89]. In the more recent field of computational photography, Veeraghavan et al. [VRA*07] showed how a 4D light field can be reconstructed from 2D sensor information by means of a coded mask. Placed at the lens, the authors achieve refocusing of images at full resolution, provided the scene being captured contains only Lambertian objects. Nayar and Mitsunaga [NM00], extended the dynamic range capabilities of an imaging system by placing a mask of spatially varying transmittance next to the sensor, and then mapping the captured information to high dynamic range.

Other works have proposed different coded apertures for defocus deblurring or depth approximation. To restore a blurred image, the apertures are designed to have a broadband frequency response, along with none (or distinguishable) zero-crossings in the Fourier domain. Hiura and Matsuyama [HM98] proposed a four-pinhole coded aperture to approximate the depth of the scene, along with a deblurred version of it, although their system required multiple images. Liang et al. [LLW*08] use a similar approach, combining tens of images captured with Hadamard-based coded patterns. Levin et al. [LFDF07] attempted to achieve all-focus and depth recovery simultaneously, relying on image statistics to design an optimal aperture and for the subsequent deconvolution. Depth recovery is limited to a multi-layered representation of the scene. Last, the idea of encoding the information before it reached the sensor has not only been limited to the spatial domain but also transferred to the temporal domain by applying a coded exposure aimed at motion deblurring [RAT06].

Another approach to recovering both a depth map of the scene and in-focus images was that of Zhou et al. [ZLN09], in this case obtaining a pair of coded apertures using both genetic algorithms and gradient descent search. The same year, a framework for evaluating coded apertures was presented, based on the quality of the resulting deblurred image and taking into account natural image statistics [ZN09]. Near-optimal apertures are obtained by means of a genetic algorithm. Recently, Masia and colleagues offered initial insights on non-binary apertures following the same approach [MCPG11], and analyzed the obtained apertures along the size, depth and shape dimensions. This paper represents a continuation of that work, which we extend by introducing two existing perceptual metrics in the optimization process leading to an aperture design, and further analyzing the potential benefits of non-binary masks.

3. The Imaging Process

Image blur due to defocus is caused by the loss of high frequency content when capturing the image. The capture process can be modeled as a convolution between the scene being captured and the point spread function (PSF) of the camera, plus some noise:

$$f = k_d * f_0 + \eta \quad (1)$$

where f_0 represents the real scene being photographed, f is the captured image, k_d is the PSF and η accounts for the noise introduced in the imaging process. Subscript d accounts for the dependency of the PSF with the defocus depth d (distance of the scene to the in-focus plane). Additionally, the PSF varies spatially across the image and depends on the absolute position of the in-focus plane as well. We will assume that the noise follows a Gaussian distribution of zero mean and a standard deviation denoted by σ , $N(0, \sigma^2)$. By means of deconvolution, an approximation \hat{f}_0 of the original sharp image can be obtained.

As Figure 1 shows, the PSF is also characterized by the pattern and size of the aperture. Since, as mentioned, blur is caused by the loss of information at certain frequencies, the response of an aperture is better analyzed in the frequency domain, where Equation 1 can be written as:

$$F = K_d \cdot F_0 + \zeta \quad (2)$$

Figure 2 shows two plots of the power spectra of different apertures: the traditional circular pattern, an optimal aperture from related previous work [ZN09], and three of the perceptually-optimized apertures presented in this paper. Note that the y-axis, showing the square of the amplitude of the response for different frequencies, is log-scale. Circular apertures exhibit zero crossings at several frequencies, and thus information at those frequencies is lost during the imaging process and cannot be recovered. Optimal apertures for deblurring therefore seek a smooth power spectrum, while keeping the transmitted energy as high as possible.



Figure 1: *Left*: Disassembled Canon EOS 50mm f/1.8 used in our tests. *Middle*: Point spread function for different apertures and degrees of defocus (*from top to bottom*: circular aperture, focused; circular aperture, defocus depth = 90cm; and one of our coded apertures, defocus depth = 80cm). *Right*: The lens with one of our coded apertures inserted.

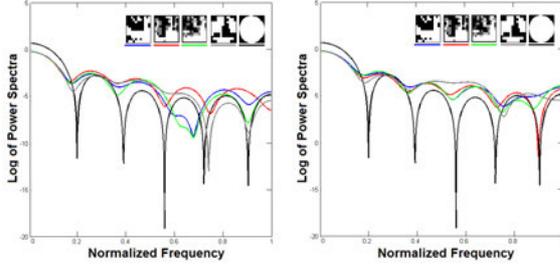


Figure 2: Power spectra of different apertures. Spectra for a conventional circular aperture and for an aperture specifically designed for defocus deblurring [ZN09] are shown in black and gray, respectively. Blue, red and green curves show the spectra of some of our perceptually-optimized apertures (please refer to the text for details).

4. Perceptual Quality Metrics

Devising an aperture pattern whose frequency response is optimal can be done in different manners. In this paper we build on the approach of Zhou and Nayar [ZN09]; in their work, the authors define their quality metric, i.e. the objective function, as the expectation of the L_2 distance between the deconvolved image \hat{F}_0 and the ground truth image F_0 with respect to ζ .

However, objective metrics working at pixel level (such as the L_2 norm) are not necessarily correlated with human perception: images with completely different per-pixel information may share a visual quality that will be easily identified by humans [Ade08]. Inspired by this observation, we introduce two additional perceptually-based metrics to guide the design of the apertures, by minimizing errors in the deconvolved images that are predicted to be perceived by a human observer. Furthermore, we include a more reliable prior based on the statistics of a large number of natural images from a recently published database [PCR10]. The perceptual metrics that we use are SSIM (Structural Similarity) [WBSS04] and the recent HDR-VDP-2 [MKRH11], which we briefly describe in the following subsections.

SSIM. The Structural Similarity Index Measure (SSIM) was introduced by Wang et al. [WBSS04], to compute the similarity between two images. It is based on a measure of structural similarity between corresponding local windows in both images. It assumes that the human visual system is very well adapted to extract structural information from a scene, and therefore evaluates the similarity between a distorted image and a reference image based on the degradation of such structural information.

Assuming x and y to be non-negative image signals, belonging to the two images to be compared, SSIM compares luminance $l(\mathbf{x}, \mathbf{y})$, contrast $c(\mathbf{x}, \mathbf{y})$, and the structure $s(\mathbf{x}, \mathbf{y})$ between the images. The latter, $s(\mathbf{x}, \mathbf{y})$, is termed structural similarity and defined as the correlation between the two im-

age signals after normalization. The three components are multiplied to obtain the final similarity measure (please refer to the original publication for details):

$$SSIM = \frac{(2\mu_x\mu_y + A_1)(2\sigma_{xy} + A_2)}{(\mu_x^2 + \mu_y^2 + A_1)(\sigma_x^2 + \sigma_y^2 + A_2)} \quad (3)$$

where μ represents mean luminance, and σ is the standard deviation, used as an estimate of the image contrast. ν is the correlation coefficient between the images, obtained as the inner product of the unit vectors $(x - \mu_x)/\sigma_x$ and $(y - \mu_y)/\sigma_y$. In our case, the local window to compute the needed statistics has been set to a 8×8 pixels square window weighted by a rotationally symmetric Gaussian function with a standard deviation $\sigma = 1.5$. The constants A_i avoid instabilities when either $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ are very close to zero; we set their values to $A_1 = (B_1L)^2$ and $A_2 = (B_2L)^2$ where L is the dynamic range of the pixel values (255 for 8-bit grayscale images), $B_1 = 0.01$, and $B_2 = 0.03$.

HDR-VDP-2. HDR-VDP-2 is a very recent metric that uses a fairly advanced model of human perception to predict both visibility of artifacts and overall quality in images [MKRH11]. The visual model used is based on existing experimental data, and accounts for all visible luminance conditions. The results of this metric show a significant improvement over its predecessor, HDR-VDP. This metric makes use of a detailed model of the optical and retinal pathway (including intra-ocular light scatter, photoreceptor spectral sensitivities and luminance masking) and takes into account contrast sensitivity for a wide range of luminances, as well as inter- and intra-channel contrast masking. We again refer the reader to the original publication for the details.

HDR-VDP-2 can yield different outputs: an estimation of the probability of detecting differences between the two images compared, or an estimation of the quality of the test image with respect to the reference image. In this work we have used the latter, a prediction of the quality degradation with respect to the reference image, expressed as a *mean-opinion-score* (from 0 to 100). We set the *color encoding* parameter of the metric to *luma-display* in order to work with the luminance channel of LDR images; the *pixels-per-degree* parameter, related to the viewing distance and the spatial resolution of the image, is set to a standard value of 30.

5. Perceptually-Optimized Apertures

The Fourier transform of the recovered image \hat{F}_0 can be obtained using Wiener deconvolution as follows [ZN09]:

$$\hat{F}_0 = \frac{F \cdot \bar{K}}{|K|^2 + |C|^2} \quad (4)$$

where \bar{K} is the complex conjugate of K , and $|K|^2 = K \cdot \bar{K}$. $|C|^2 = C \cdot \bar{C}$ is the matrix of noise-to-signal power ratios

(NSR) of the additive noise. We precompute this matrix as $|C|^2 = \sigma^2/S$, where S is the estimated power spectra of a natural image and σ^2 is the noise variance. To estimate S , we rely on recent work on statistics of natural images by Pouli et al. [PCR10], and select from their database 180 images from an extensive collection of two different categories: half of the images belong to the *manmade-outdoors* category, while the other half belongs to the *natural* category. The estimated power spectra is obtained as the average of the power spectra over small windows of each of the 180 images and will be used as our prior in the deconvolution process.

The quality of the recovered image \hat{f}_0 with respect to the real image f_0 is measured using a combination of the $L2$ norm, the *SSIM* index and the *HDR-VDP-2* score (*VDP2*). The aperture quality metric Q is then given by:

$$Q = \lambda_1(1 - L2) + \lambda_2(SSIM) + \lambda_3(VDP2/100) \quad (5)$$

For the normalized $L2$ norm, 0 represents perfect quality, while 1 means worst quality. The *SSIM* index can yield values in the range $[-1, 1]$, but we observe that for the specific case of blurred images the structural information does not change enough for the index to reach negative values. Therefore, values for the *SSIM* index range from 0 (worst quality) to 1 (best quality). The values for *VDP2* range from 0 (worst quality) to 100 (best quality). Last, the vector $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ represents the weights assigned to each metric (discussed in Subsection 5.1).

5.1. Optimization

Our goal is to obtain apertures with the largest possible Q value according to our quality metric. Once we have introduced a way of evaluating a certain aperture with Equation 5, an optimization method can be used to obtain the maximum value of Q over the space of all possible apertures. This space is infinite, limited only by physical restrictions (i.e. apertures with negative values are not realizable in practice and resolution is limited by the printing process). Resolution is additionally limited by diffraction effects, which appear as the size of the pixels in the aperture gets smaller, and hinder its performance. In our case, we fix the resolution of the apertures to 11×11 .

Transmissivity is an additional issue to be taken into account when designing an aperture. Coded apertures typically have lower transmission rates than their circular counterparts, and the use of a longer exposure time to obtain an equivalent brightness to that of the circular aperture can cause other problems such as motion blur. We fix the transmission rate in our apertures to 0.578. We have chosen this value empirically since it yields adequate exposure times, while being similar to other coded apertures proposed for defocused deblurring.

In order to search for the best aperture pattern we

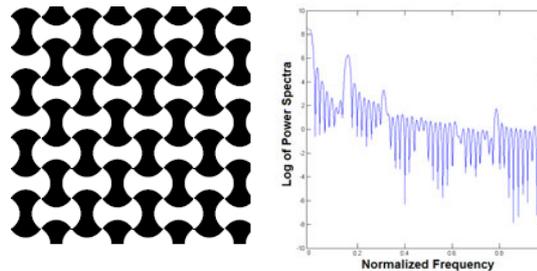


Figure 3: *Left*: Image pattern, after [JSK08], used in the evaluation function of the genetic algorithm. *Right*: Wide bandwidth power spectra of the selected pattern.

have implemented a genetic algorithm (similar to [ZN09, MCPG11]), which uses our novel quality metric as evaluation function (i.e. objective function). The algorithm has the following scheme:

- *Initialization*. An initial population of $N = 1500$ apertures is randomly generated. An aperture is defined by a vector of $P = 121$ elements, each element corresponding to an aperture pixel.
- *Selection*. We evaluate each aperture by simulating the capture process, multiplying the Fourier transform of a sharp image F_0 by the OTF (response of the aperture in the frequency domain) and adding the Fourier transform of the gaussian noise (Equation 2). We then perform Wiener deconvolution with our prior $|C|^2$ of natural images (Equation 4). The quality of the recovered image is measured using our quality metric Q (Equation 5), and the $M = 150$ apertures with best quality result are selected. The image used to perform this step, which is 200×200 pixels in size, is similar to the pattern used by Joshi et al. [JSK08] (see Figure 3), since this pattern has a wide bandwidth spectrum in the frequency domain.
- *Reproduction*. The M selected apertures are used to populate the next generation by means of crossover and mutation. Crossover implies randomly selecting two apertures, duplicating them, and exchanging corresponding bits between them with probability $c_1 = 0.2$, obtaining two new apertures. Mutation ensures diversity by modifying each bit of the aperture with probability $c_2 = 0.05$.
- *Termination*. The reproduction and selection steps are repeated until the termination condition is met. In our case, the algorithm stops when the increment in the quality factor is less than 0.1%, which generally occurs before $G = 80$ generations.

The standard deviation of the noise applied in the *selection* process is set in principle to $\sigma = 0.005$ (we later explore this parameter in Section 6.2). This is based on previous findings where apertures designed for σ values of 0.001 and 0.005 proved to work best for a wide variety of images [MCPG11]. Following Equation 5, we consider four

variations of our evaluation function, characterized by the weight assigned to each metric:

- $\Lambda = \{1, 0, 0\}$: just using the $L2$ norm
- $\Lambda = \{0, 1, 0\}$: just $SSIM$
- $\Lambda = \{0, 0, 1\}$: just $HDR-VDP-2$
- $\Lambda = \{1, 1, 1\}$: combining $L2$, $SSIM$, and $HDR-VDP-2$

We have run the genetic algorithm three times for each variation of the evaluation function, yielding three executions to which we will refer as $I = \{1, 2, 3\}$. The top row for each weight vector Λ in Figure 4 shows the twelve binary apertures obtained. The other two rows show the results for non-binary apertures, which will be discussed in Section 7.

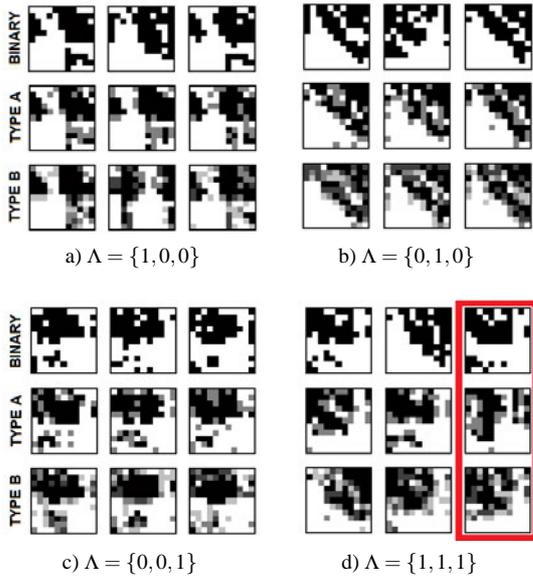


Figure 4: Apertures obtained for the four variations of the evaluation function. For each weight vector Λ , the top row shows the results of the binary apertures; while second and third rows show the non-binary type A and non-binary type B results (see Section 7). Columns correspond to the different executions $I = \{1, 2, 3\}$. The apertures which exhibit the best performance (Section 6) are highlighted in red.

6. Performance of the Apertures

In this section, we restrict the analysis of performance to binary apertures; non-binary apertures will be discussed in Section 7. We simulate the capture process by first convolving a sharp image f_0 with the aperture k_d and adding noise η as described by Equation 1. To recover the deblurred image \hat{f}_0 , we perform Wiener deconvolution using our prior $|C|^2$ derived from natural images (Equation 4). Note that in practice we work in the frequency domain.

The quality of each recovered image is measured using the $L2$ norm, the SSIM index and the HDR-VDP-2 score.

In order to take in account the results of all three metrics together we calculate the *aggregate* quality factor Q_a as:

$$Q_a = (1 - L2) + (SSIM) + (VDP2/100) \quad (6)$$

where larger values of Q_a correspond to better quality in the recovered images ($Q_a \in [0, 3]$).



Figure 5: Some of the images used for evaluating the obtained apertures. Image licensed under Creative Commons copyright of freemages and flickr users (in reading order) Christophe Eyquem, Stig Nygaard, Paola Farrera and Juampe Lopez.

We repeat this process using 30 images of different types of scenes (nature, people, buildings), in order to include a large and varied enough selection. A few examples of the images used are shown in Figure 5. For each aperture, we calculate the values for the three different metrics plus the aggregate quality factor Q_a for the 30 recovered images. We therefore have, for each type of aperture (binary, type A or type B) and each weight vector Λ , a total of 90 Q_a values. We denote each of these values as $Q_a(i, j)$, where i refers to the execution number ($I = \{1, 2, 3\}$) and j to the image number ($J = [1..30]$)[†]. In the following we analyze separately the influence of the perceptual metrics and the noise level in the performance of the obtained apertures.

6.1. Influence of the Perceptual Metrics

We compute the aggregate quality factor of the *best binary* aperture obtained for each Λ averaged along the 30 images $Q_a(i_{best}, J)$ (together with the corresponding standard deviation); we also compute the mean along the 30 test images of

[†] Note that Q_a values conform a four-dimensional set of data. One dimension corresponds to the type of aperture (binary, type A, or type B), another dimension is the weight vector Λ , and the third and fourth dimensions are the number of executions $i \in I$ and the number of test images $j \in J$.

the individual scores of the three metrics $L2$, $SSIM$ and $HDR-VDP-2$. These serve as an indicative of the performance of a particular aperture. Additionally, we obtain the mean aggregate quality factor of the three executions, $Q_{a(I,J)}$, together with its standard deviation $\sigma(Q_{a(I,J)})$. These values will illustrate the appropriateness of including each of the perceptual metrics in the evaluation function.

Table 1 compiles these results for binary apertures. The first five columns refer to individual data for the best aperture of the three executions, whereas the last two refer to the averaged values for that particular evaluation function:

$$Q_{a(I,J)} = \frac{1}{|I|} \sum_i \left(\frac{1}{|J|} \sum_j Q_{a(i,j)} \right), \quad (7)$$

with $|I| = 3$ and $|J| = 30$. It can be seen how the combination of the three metrics ($\Lambda = \{1, 1, 1\}$) yields the highest Q_a scores, which translates into better apertures for defocus deblurring. Although we have limited ourselves in this paper to equal weights when combining the three metrics, leaving further exploration of other possibilities for future work, these results clearly suggest the benefits of using perceptual metrics when deriving the apertures.

6.2. Influence of Noise

The apertures analyzed so far have all been computed assuming an image noise level of $\sigma = 0.005$. We now explore performance of our apertures over a wider range of noise levels, to ensure that our findings generalize to different image conditions. Figure 6 shows $L2$, $SSIM$, $HDR-VDP-2$ and Q_a for images captured and deblurred using our best perceptually-optimized binary aperture. The images used are the same 30 test images described before, but after synthetically adding to them noise of increasing standard deviation: $\sigma = 0.0001, 0.0005, 0.001, 0.002, 0.005, 0.008, 0.01$ and 0.02 . It can be seen how our optimized patterns perform well across all noise levels, in contrast to standard circular apertures which have been proved to be very sensitive to high noise levels [ZN09].

6.3. Comparison with Other Metrics

We now compare the performance of our best binary aperture (marked in red in Figure 4) with a conventional circular aperture and with the best aperture described by Zhou et al. [ZN09] for a noise level of $\sigma = 0.005$. Note that Zhou's aperture has been optimized using only a $L2$ norm quality metric.

Figure 7 shows the results for both comparisons (top: against a circular aperture; bottom: against Zhou's aperture). We have used each of the three metrics to compare the quality of corresponding recovered images. Each dot in the diagrams represents the values obtained for a given image in

the 30-image data set used in this paper. Thus, values on the diagonal would indicate equal performance of the two apertures being compared. For the case of the $L2$ norm, values above the diagonal favor our binary aperture (plotted in the x-axis), whereas for the other two metrics, values below the diagonal are preferred. It is clear from these data that our binary aperture consistently outperforms not only the conventional circular aperture, but Zhou's aperture as well (although obviously by a lesser margin). This translates into recovered images of better quality according to all the metrics, as will be shown in Section 8.

7. Non-Binary Apertures

Binary codes have the initial advantage of reducing the search space, and are usually preferred in the existing literature. However, there is no principled motivation to restrict the aperture pixel values to either black or white, other than apparent simplicity. A notable exception in this regard is the work by Veeraraghavan and colleagues [VRA*07], where the authors report the advantages of continuous-valued apertures, found by gradient descent optimization: reduced computational times and less noise in the recovered (deblurred) images.

In order to analyze if our perceptual metrics also improve the performance of non-binary apertures, we repeat our optimization process, but allowing the solutions of the genetic algorithm to include values between 0 and 1. In order to limit the search space, in practice we restrict the set of possible values to i) one level of gray (the allowed pixel values thus being $\{0, 0.5, 1\}$) and ii) three levels of gray ($\{0, 0.25, 0.5, 0.75, 1\}$). We call the results of both options non-binary type A and non-binary type B, respectively. The middle and bottom rows in Figure 4 show the apertures obtained for both types (again, we obtain three different apertures for each weight vector Λ).

We perform the same simulated validation as described in Section 6 for the binary apertures. Our results confirm that again the combination of the three metrics with equal weights $\Lambda = \{1, 1, 1\}$ yields apertures with better overall performance. Table 2 summarizes the results. In an analogous manner to the analysis for binary apertures, the first five columns show data for the best non-binary aperture in each case, averaged across the 30 test images. The last two columns show averaged values across the 30 images and the three executions computed for each evaluation function.

8. Results and Discussion

While in the previous sections we have evaluated the performance of the apertures by simulating the capture process, in this section we test our apertures on a real scenario; we print and insert the masks into a camera, calibrate the system, and capture real scenes. We have used a Canon EOS 500D with a EF 50mm f/1.8 II lens, shown (disassembled) in Figure 1.

Binary							
	$L2(i_{best}, J)$	$SSIM(i_{best}, J)$	$VDP2(i_{best}, J)$	$Q_a(i_{best}, J)$	$\sigma(Q_a(i_{best}, J))$	$Q_a(I, J)$	$\sigma(Q_a(I, J))$
$\Lambda = \{1, 0, 0\}$	1.0893	0.8870	68.6038	2.5622	0.1405	2.5564	0.0042
$\Lambda = \{0, 1, 0\}$	1.1716	0.8994	71.7686	2.6054	0.1164	2.5479	0.0407
$\Lambda = \{0, 0, 1\}$	1.0359	0.8883	70.9459	2.5874	0.1190	2.5619	0.0183
$\Lambda = \{1, 1, 1\}$	1.0261	0.8990	74.4169	2.6329	0.1026	2.5921	0.0315

Table 1: Performance evaluation of binary apertures obtained with the different objective functions (i.e. different weight vector Λ). The first five columns of each table show values of the different metrics and aggregate quality factor for the best binary apertures of each evaluation function averaged across the 30 test images, plus the standard deviation of the latter. The two rightmost columns show, for each evaluation function, the mean aggregate quality factor of the three executions and its standard deviation. Note that the L2 norm is shown as a percentage with respect to the maximum error.

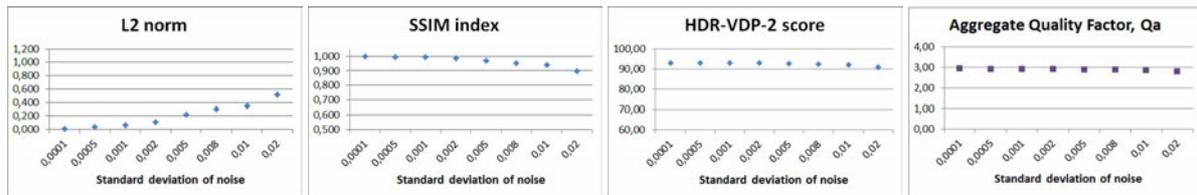


Figure 6: Performance of the best perceptually-optimized binary coded aperture across eight different levels of noise, measured with the L2, SSIM, HDR-VDP-2 and Q_a metrics. The L2 norm shows percentages with respect to the maximum error.

To calibrate the response of the camera (PSF) at different depths, we used a LED which we made as close as possible to a point light source with the aid of a pierced thick black cardboard. We locked the focus at 1,20 m and took an initial focused image, followed by images of the LED at 20, 40, 60 and 80 cm with respect to the in-focus plane. For each depth, the actual cropped image of the LED served us as PSF, after appropriate thresholding of surrounding values which contain residual light, and subsequent normalization for energy conservation purposes. The resulting PSFs for one of our binary apertures are shown in Figure 8, next to the PSFs of a conventional, circular aperture for comparison purposes.

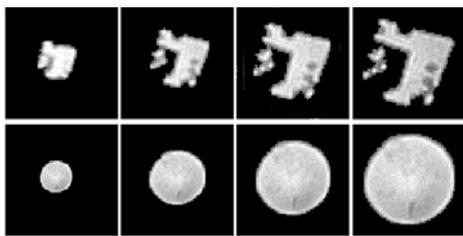


Figure 8: PSFs at four different defocus depths (20, 40, 60 and 80 cm). *Top row*: For our binary coded aperture. *Bottom row*: For a circular aperture.

Once calibration has been performed, images of three scenes at the four defocus depths (20, 40, 60 and 80 cm) were taken with each of the selected apertures. During the capture

process, the aperture was set to F2.0, and the exposure time to 1/20 for all scenes and apertures, to ensure a fair comparison. The captured defocused images are then deblurred using the corresponding calibrated PSF by means of Wiener deconvolution. We used Wiener deconvolution with a NSR of 0.005 instead of the prior of natural images, since in real experiments it gave better results. This may be caused by the fact that our prior $|C|^2$ is calculated with the power spectra of images from *manmade day* and *natural day* scenes, which have similar spectral slopes, while the spectral slope for images from *manmade indoors* scenes (similar to the scenes we capture) is slightly different [PCR10]. The same exposure and aperture settings were used for all our coded apertures. Figure 9 depicts the recovered images for three different apertures: a circular aperture, our best binary coded aperture and the best aperture obtained by Zhou et al. [ZN09] for a noise value of $\sigma = 0.005$, to which we have also compared in Section 6. Defocus depths are 60 cm for recovered images (b), (c) and (d) and 80 cm for (e) and (f). Insets depict the corresponding PSF.

Our aperture clearly outperforms the circular one, which was to be expected from the existing body of literature about coded apertures. More interesting is the comparison with a current state-of-the-art coded aperture; when compared to the aperture described by Zhou et al., our perceptually-optimized approach yields less ringing artifacts, exhibiting, qualitatively, a better overall performance. Additional results for two other scenes at four defocus depths (20, 40, 60 and 80 cm) can be seen in Figure 10. Please note that the slight

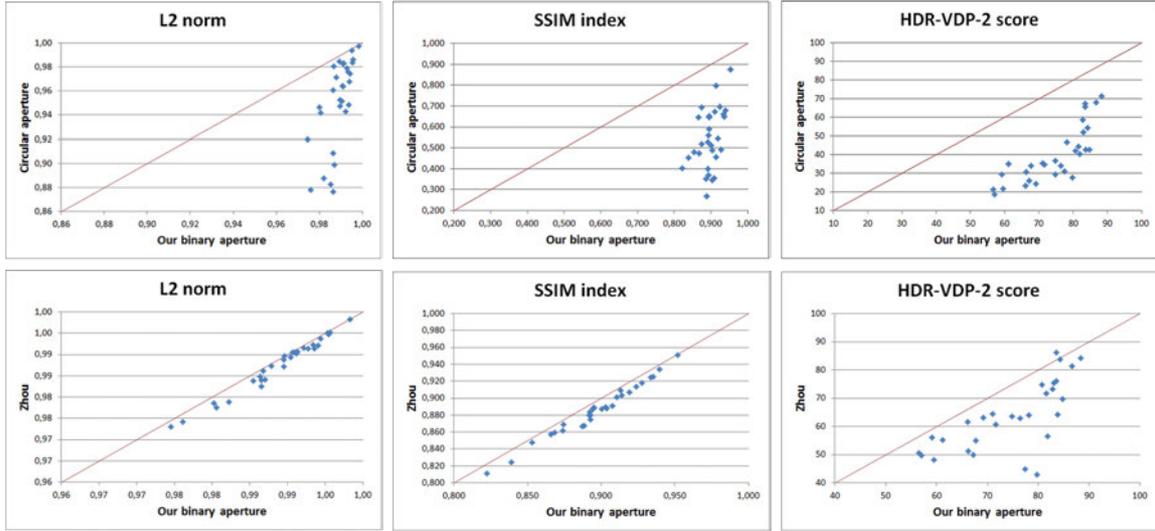


Figure 7: Scatter plots showing the performance of our best binary coded aperture against that of a circular aperture (*top row*) and against the coded aperture proposed by Zhou et al. [ZNO9] for an image noise of $\sigma = 0.005$ (*bottom row*). For the sake of consistency, the $L2$ norm is depicted as $(1 - L2/100)$, $L2$ being the percentage with respect to the maximum error. It can be seen how our proposed aperture outperforms the other two.

Non-binary type A							
	$L2(i_{best}, J)$	$SSIM(i_{best}, J)$	$VDP2(i_{best}, J)$	$Q_a(i_{best}, J)$	$\sigma(Q_a(i_{best}, J))$	$Q_a(I, J)$	$\sigma(Q_a(I, J))$
$\Lambda = \{1, 0, 0\}$	1.0671	0.8887	69.3253	2.5812	0.1265	2.5645	0.0053
$\Lambda = \{0, 1, 0\}$	1.1840	0.9012	63.4331	2.5245	0.1435	2.5258	0.0026
$\Lambda = \{0, 0, 1\}$	1.0707	0.8867	70.0342	2.5763	0.1352	2.5584	0.0127
$\Lambda = \{1, 1, 1\}$	1.1368	0.8928	72.3652	2.6050	0.1186	2.5963	0.0069

Non-binary type B							
	$L2(i_{best}, J)$	$SSIM(i_{best}, J)$	$VDP2(i_{best}, J)$	$Q_a(i_{best}, J)$	$\sigma(Q_a(i_{best}, J))$	$Q_a(I, J)$	$\sigma(Q_a(I, J))$
$\Lambda = \{1, 0, 0\}$	1.1436	0.8880	68.9516	2.5660	0.1460	2.5460	0.0168
$\Lambda = \{0, 1, 0\}$	1.1979	0.9022	63.1623	2.5218	0.1399	2.5153	0.0054
$\Lambda = \{0, 0, 1\}$	1.1392	0.8869	69.4941	2.5705	0.1203	2.5460	0.0019
$\Lambda = \{1, 1, 1\}$	1.1542	0.8998	70.8276	2.5965	0.1224	2.5692	0.0234

Table 2: Performance evaluation of non-binary apertures obtained with the different objective functions (i.e. different weight vector Λ). The first five columns show values of the different metrics and aggregate quality factor for the best non-binary apertures of each evaluation function averaged across the 30 test images, plus the standard deviation of the latter. The two rightmost columns show, for each evaluation function, the mean aggregate quality factor of the three executions and its standard deviation. Note that the $L2$ norm is shown as a percentage with respect to the maximum error.

changes in brightness in the images are due to different illumination conditions, and not to the light transmitted by the aperture.

Minor artifacts that appear in our recovered images are probably due to errors in the calibrated PSF. Another possible cause of error may be inaccurately modeled image noise [SJA08]. Additionally, although the PSF actually varies spatially across the image [LFDf07], we consider here one sin-

gle PSF, measured at the center of the image, for the entire image plane.

The non-binary apertures obtained in Section 7 were also evaluated in a real scenario. Figure 12 shows the recovered images obtained with the best binary aperture (left), the best non-binary aperture of type A (middle) and the best non-binary aperture of type B (right). Although non-binary apertures seem to yield images with lower background noise, evi-

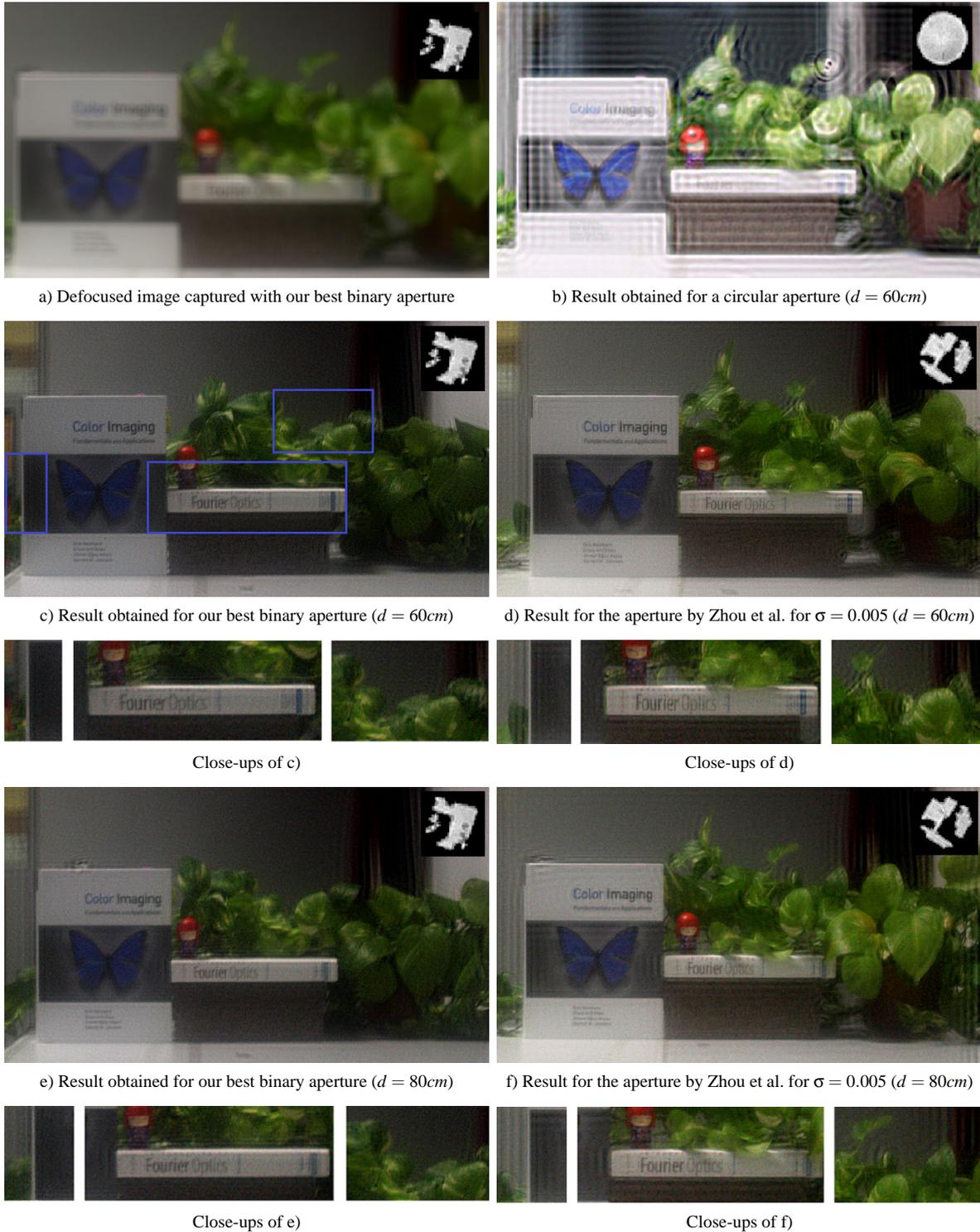


Figure 9: Recovered images for different apertures (circular, Zhou's for $\sigma = 0.005$ and our best perceptually-optimized binary aperture) and different defocus depths d . Close-ups of this images show the improved quality and fewer ringing artifacts of images recovered with the perceptually-optimized aperture. Insets depict the PSF of the aperture used in each case. Note that results for the circular aperture are significantly brighter because of its higher transmission rate.

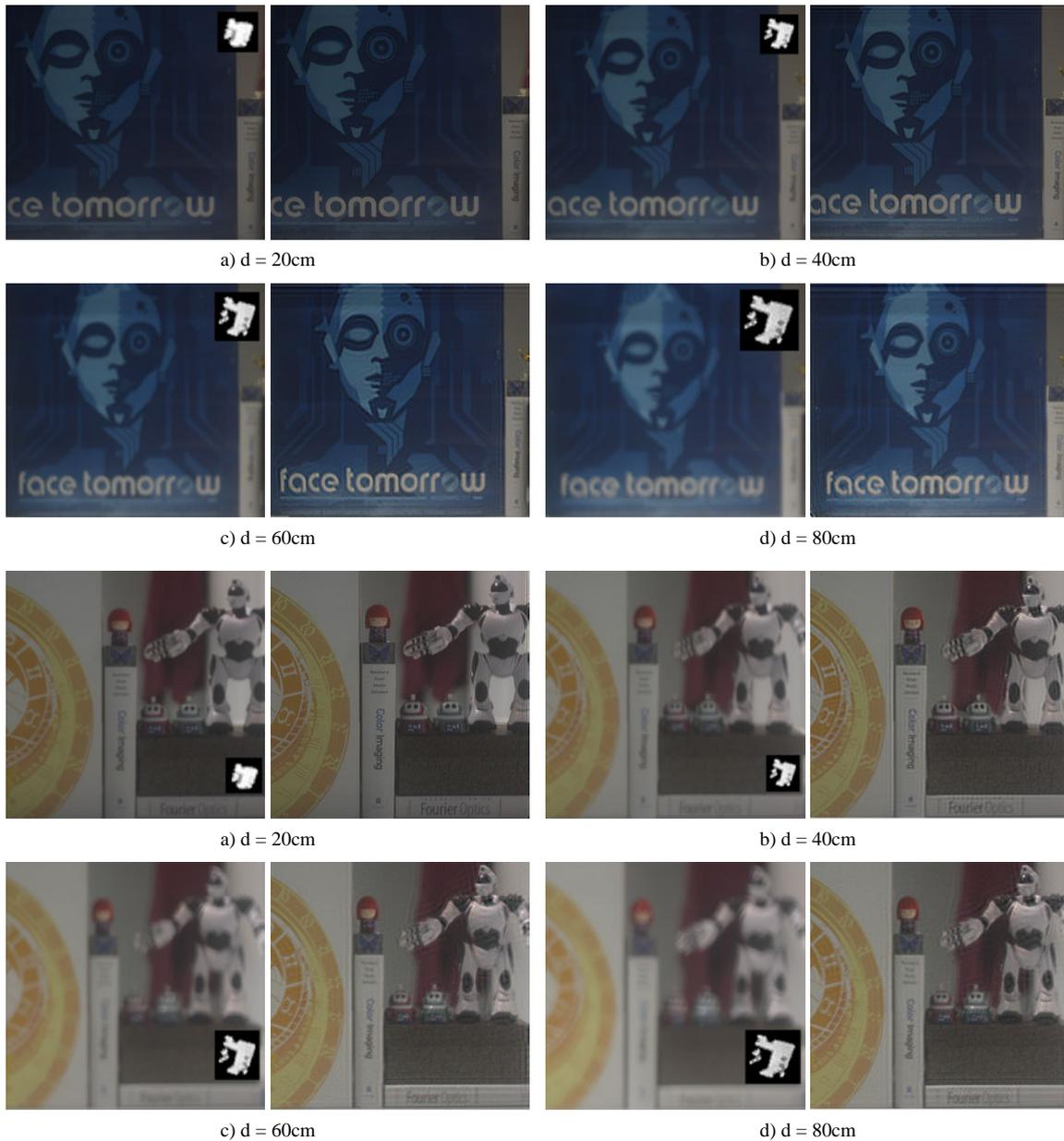


Figure 10: Defocused and recovered images at four different defocus depths d obtained with the perceptually-optimized binary coded aperture for two different scenes.

dence is not strong enough to derive any definite conclusion. It is worth noting that metrics based on simulations of the capture process yield similar quality values for binary apertures and their non-binary counterparts (see Tables 1 and 2). This may suggest the need for a more complex image formation model, essentially in what regards to the additive noise, a need which has already been observed by other authors in the field [VRA*07].

Observations from real-world images are consistent

with the power spectra shown in Figure 2, where our perceptually-optimized apertures exhibit larger amplitudes for the majority of the spectrum compared to Zhou's and the circular aperture. Additionally, in order to assess how well real results correlate with simulated ones we have compared results from a real setup with results simulated for the same conditions. We have done this for our best binary coded aperture selected in red in Figure 4. To do this we compute the size of the blur for the different defocus depths used in

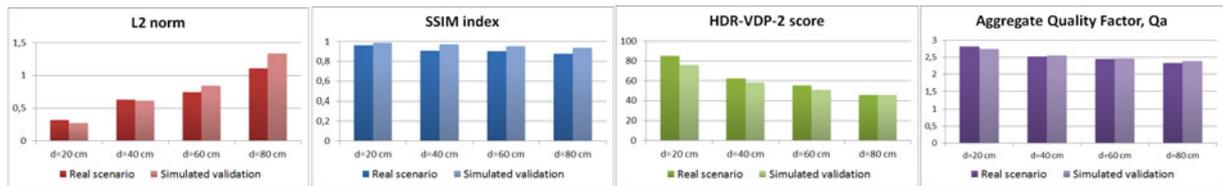


Figure 11: Correlation between real-capture and simulated-capture results. Average quality of the recovered images for both cases (real and simulated) according to each metric for the four defocus depths tested (20, 40, 60 and 80 cm) and to the *aggregate* quality factor Q_a calculated according to Equation 6.

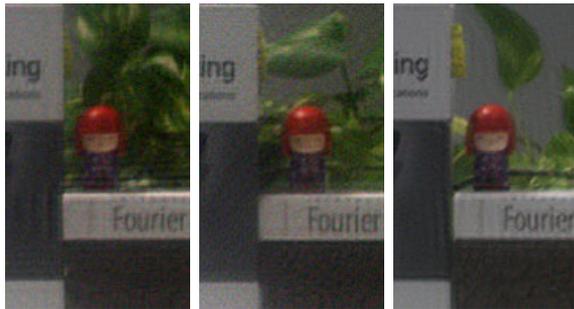


Figure 12: Comparison between deblurred images captured using perceptually-optimized binary (*left*), non-binary type A (*middle*), and non-binary type B (*right*) apertures.

the real scenario (20, 40, 60 and 80 cm) and scale the PSF accordingly when computing the simulated blurred images. Although this scaling is only an approximation to what the real PSF would be, it does give information on how well simulated results extrapolate to real results. Figure 11 shows the results obtained by the different quality metrics (plus the aggregate factor Q_a) for real and simulated results. We can clearly see how both exhibit the same behavior and trends, thus showing the validity of the use of simulated capture processes for the evaluation of the different apertures.

Finally, the time until convergence when running the algorithm on an Intel core i7 930 @2.80GHz is 13,72 hours for the evaluation function using $(\Lambda = 1, 1, 1)$, which is obviously the most expensive scenario. As expected, computing the HDR-VDP 2 metric consumes the largest amount of time (62% of the total execution time when $\Lambda = 1, 1, 1$), followed by SSIM; there is clearly a trade-off between complexity of the metrics included and performance of the resulting apertures.

9. Conclusions and Future Work

In this paper we have presented a method to obtain coded apertures for defocus deblurring, which takes into account human perception for the computation of the optimal aperture pattern. Following previous approaches, we pose the problem as an optimization, and, to our knowledge, propose

the first algorithm that makes use of perceptual quality metrics in the objective function. We explore the performance of different quality metrics for the design of coded apertures, including the well-established SSIM, and the state-of-the-art HDR-VDP-2, which features a comprehensive model of the HVS, as well as the $L2$ norm, previously used in related works. The results obtained show that the best apertures are obtained when a combination of the three metrics is used in the objective function, clearly outperforms existing apertures, both in simulated and real scenarios, results obtained by conventional circular apertures and by an existing aperture pattern specifically designed for defocus deblurring.

Additionally, we have explored non-binary aperture patterns, often neglected in the literature. Even though results with real images seem to indicate a better performance (i.e. less ringing artifacts) of non-binary apertures with respect to their binary counterparts, sharpness appears somewhat hindered by non-binary masks in comparison to binary patterns, resulting in a trade-off between both.

The most important challenge for the future is probably devising a new model for the noise inherent to the capture process, which would allow a better modeling of the process and thus a better design of coded aperture patterns. Although we show that simulated and real results correlate fairly well, differences remain, which may be overcome with a better model.

10. Acknowledgements

We would like to thank the reviewers for their valuable comments. We also thank Changyin Zhou for his insights, and Javier Marco for his assistance during the capture sessions. We will also like to thank freemages and flickr users Christophe Eyquem, Stig Nygaard, Juampe Lopez and Paola Farrera. This research has been funded by the European Commission, Seventh Framework Programme, through the projects GOLEM (Marie Curie IAPP, grant agreement no.: 251415) and VERVE (Information and Communication Technologies, grant agreement no.: 288914), and by the Spanish Ministry of Science and Technology (TIN2010-21543). Belen Masia is supported by a FPU grant from the Spanish Ministry of Education.

References

- [Ade08] ADELSON E. H.: Image statistics and surface perception. In *Human Vision and Electronic Imaging XIII, Proceedings of the SPIE* (2008), no. 1, SPIE. 3
- [GF89] GOTTESMAN S., FENIMORE E.: New family of binary arrays for coded aperture imaging. *Applied Optics*, 20 (1989), 4344–4352. 2
- [HM98] HIURA S., MATSUYAMA T.: Depth measurement by the multi-focus camera. In *IEEE Conference on Computer Vision and Pattern Recognition* (Washington DC, USA, 1998), IEEE Computer Society. 2
- [ItZ92] INTZAND J.: *Coded Aperture Imaging in High-Energy Astronomy*. PhD thesis, University of Utrecht, 1992. 2
- [JSK08] JOSHI N., SZELISKI R., KRIEGMAN D. J.: PSF Estimation Using Sharp Edge Prediction. In *IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage, Alaska, USA, 2008), IEEE Computer Society. 4
- [LFD07] LEVIN A., FERGUS R., DURAND F., FREEMAN W.: Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics* 26, 3 (2007). 2, 8
- [LLW*08] LIANG C., LIN T., WONG B., LIU C., CHEN H.: Programmable aperture photography: multiplexed light field acquisition. *ACM Transactions on Graphics* 27, 3 (2008). 2
- [MCPG11] MASIA B., CORRALES A., PRESA L., GUTIERREZ D.: Coded apertures for defocus deblurring. In *Symposium Iberoamericano de Computacion Grafica* (Faro, Portugal, 2011). 2, 4
- [MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics* 30, 40 (2011). 3
- [NM00] NAYAR S., MITSUNAGA T.: High dynamic range imaging: spatially varying pixel exposures. In *IEEE Conference on Computer Vision and Pattern Recognition* (Hilton Head, SC, USA, 2000), IEEE Computer Society, pp. 472–479. 2
- [PCR10] POULI T., CUNNINGHAM D., REINHARD E.: Statistical regularities in low and high dynamic range images. *ACM Symposium on Applied Perception in Graphics and Visualization (APGV)* (July 2010). 3, 4, 7
- [RAT06] RASKAR R., AGRAWAL A., TUMBLIN J.: Coded exposure photography: Motion deblurring using shuttered camera. *ACM Transactions on Graphics* 25, 3 (2006), 795–804. 2
- [SJA08] SHAN Q., JIA J., AGARWALA A.: High-quality Motion Deblurring from a Single Image. *ACM Transactions on Graphics* 27, 3 (August 2008). 8
- [VRA*07] VEERARAGHAVAN A., RASKAR R., AGRAWAL A., MOHAN A., TUMBLIN J.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics* 26 (July 2007). 2, 6, 10
- [WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (April 2004). 3
- [ZLN09] ZHOU C., LIN S., NAYAR S.: Coded aperture pairs for depth from defocus. In *IEEE International Conference on Computer Vision (ICCV)* (Kyoto, Japan, 2009). 2
- [ZN09] ZHOU C., NAYAR S. K.: What are Good Apertures for Defocus Deblurring? In *IEEE International Conference on Computational Photography* (San Francisco, CA, USA, 2009). 1, 2, 3, 4, 6, 7, 8