Optimization-Based Adaptive Optical Correction for Holographic Projectors

Andrzej Kaczorowski, George S. Gordon, Ananta Palani, Stanisław Czerniawski, and Timothy D. Wilkinson

***Abstract*—A novel method of correcting the aberrations of holographic projectors is presented. The method employs an optimization algorithm to determine an aberration-correcting phase mask composed of 13 Zernike Polynomials. The mask can be used thereafter to correct every image produced by the projector. Two optimization algorithms are demonstrated: a hybrid genetic steepest descent algorithm and a heuristic variant of steepest descent. The primary advantage of these methods is that no modifications of the standard holographic projector are required. Furthermore, the methods are fully automated. They are evaluated on two projectors with different Spatial Light Modulator flatness profiles for three wavelengths. First, the correction is demonstrated for both projectors on a green wavelength. It is then adapted for red and blue wavelengths by rescaling the mask and adjusting for chromatic aberration. The hybrid genetic-steepest-descent algorithm is compared with the heuristic steepest descent algorithm. On average, the hybrid algorithm is found to give better and more reliable correction than the heuristic steepest descent algorithm while taking 50% longer to terminate. The method is also compared with non-automated interferometric flatness measurements and is found to produce improved results.**

***Index Terms*—Adaptive optics, holography, spatial light modulators (SLMs).**

I. Introduction

SPATIAL LIGHT modulators (SLMs) have been successfully deployed for use in conventional projectors as imaging devices [1]. However, they are also extensively used in a variety of applications, including optical correlators [2], [3], fiber-optic systems [4], [5], and Fourier projectors [1], [6]–[8], which rely on their ability to operate in the Fourier plane and display holograms. Holographic imaging systems present numerous advantages over conventional ones, such as high efficiency, potential for miniaturization, and aberration correction [1], [6], [7]. Diffraction is the basis of holography and requires a high spatial and temporal coherence of light to achieve high-quality performance.

Manuscript received November 28, 2014; revised February 19, 2015; accepted March 24, 2015. Date of publication April 02, 2015; date of current version July 13, 2015. This work was supported by Engineering and Physical Sciences Research Council. The work of G. S. Gordon was supported by Henslow Research Fellowship, U.K.

A. Kaczorowski, G. S. Gordon, A. Palani, and T. D. Wilkinson are with the Electrical Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: ak762@cam.ac.uk; gsdg2@cam.ac.uk; ap648@cam.ac.uk; tdw@eng.cam.ac.uk).

S. Czerniawski was with Department of Computing, Imperial College, London SW7 2AZ, U.K. He is now with Czerniawski Consulting Ltd., Cambridge CB1 9ER, U.K. (e-mail: czer.consulting@gmail.com).

Digital Object Identifier 10.1109/JDT.2015.2418436

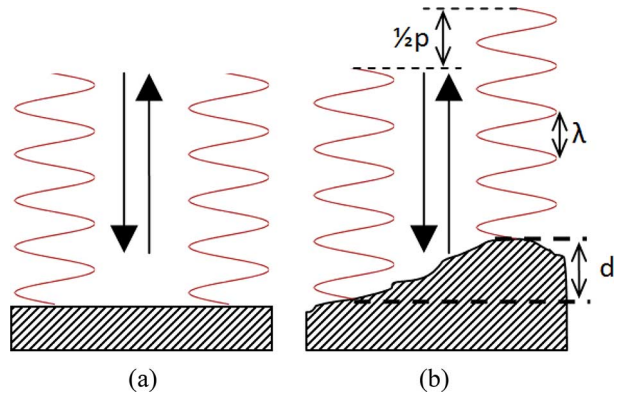


Fig. 1. Effect of non-flatness on a coherent beam of light: (a) flat surface, (b) non-flat surface.

Optical aberrations of any sort introduce unwanted spatially varying phase changes that degrade images produced using coherent light. In systems using SLMs, the non-flatness of the surface is usually the most significant cause of aberration, given the ready availability of high-quality static optical elements (lenses, mirrors etc.). Conventional imaging systems typically use incoherent illumination and, as a result, they have a high tolerance for flatness errors on the order of few micrometers [1]. By contrast, holographic imaging systems, such as Fourier projectors, have a more stringent requirement for flat optical surfaces [9].

When light reflects off a flat surface, the wavefront is preserved, as shown in Fig. 1(a). When light reflects of a surface that is not flat as shown in Fig. 1(b), its phase changes by a factor of $p = 2\pi\frac{2d}{\lambda}$ radians, where d is the difference in flatness and λ is the wavelength of light. Assuming that the difference in flatness is 2 μm (a large, but still acceptable value in conventional projectors [1]) and the wavelength of light is 500 nm, the phase change would be 16π radians—eight wavelengths.

This is a substantial error that degrades the spatial coherence of a beam, and will severely reduce the quality of an image. In order to produce an image of high quality, the deviation in flatness should be substantially smaller than the wavelength of light. A common rule of thumb suggests that the peak-valley aberration should be less than a quarter of the wavelength (Rayleigh's quarter wavelength rule) [9].

This problem has been addressed previously using an interferometer to measure the interference fringes of light reflected from the SLM to characterize the surface non-flatness [1]. However, this method has a number of disadvantages—most notably, it requires the use of a separate experiment, which may have different aberrations that the ones present in the actual Fourier

Projector. Furthermore, the fringe patterns must be manually decoded into a phase mask, which is a lengthy procedure.

Several other systems have successfully used deformable mirrors (DMs) to correct for non-flatness and other aberrations [10], [11]. For a Fourier projector, this approach would require incorporating a deformable mirror into the setup, which would add cost and complexity.

We present a novel way to characterize the aberrations of a Fourier projector using a feedback-loop mechanism. In this work, the flatness is not measured directly, but instead, its effects on the image, in particular the physical size of a single effective pixel, are measured and characterized. This incorporates not only the non-flatness of the SLM but also the aberrations of other components in the setup such as the laser, lenses, beamsplitters, and mirrors.

First, a brief theoretical discussion of aberration correction using holograms is presented. Next, the method of judging the amount of aberration present in the image is introduced. This provides an accurate measure of the quality of a spot displayed on the screen, despite effects such as saturation and limited bit-depth often found in low-cost webcams.

Next, it is necessary to determine a corrective phase mask comprising a weighted sum of Zernike polynomials for which the value of the fit function is the smallest. This is formulated as an optimization problem and two algorithms are considered: a heuristic variant of steepest descent and a hybrid genetic-steepest descent algorithm.

Finally, the different optimization algorithms are compared. Heuristic Steepest Descent is found to perform quite well, but suffers from a tendency to become stuck in local minima, i.e., a correction that is not ideal, but is better than all the possibilities nearby. The hybrid genetic-steepest-descent is found to converge more reliably, because it explores the correction space in pseudo-random manner. However, it is found to take 50% longer to converge than the heuristic steepest descent.

II. Theoretical Background

A. Aberrations in Holography

The replay field of an image in the presence of aberrations can be represented as [4]

$$\psi(u,v) = \mathcal{F}\left[\mathrm{B}(x,y)\mathrm{H}(x,y)\mathrm{e}^{i\varphi(x,y)}\right] \tag{1}$$

where $\psi(u,v)$ is the replay field, $H(x,y)$ is the hologram, $B(x,y)$ is the illumination profile of a laser, $\varphi(x,y)$ is the phase profile (aberrations) and $\mathcal{F}[]$ denotes the Fourier Transform operation.

In this work, it is assumed that the illumination is uniform.

In order to correct the distorted replay field, a corrective phase profile must be determined that compensates $\varphi(x,y)$. This corrective profile can be approximated as a linear combination of Zernike Polynomials.

Zernike Polynomials are widely used in optics, because they correspond to common aberrations, such as defocus, astigmatism, coma, spherical aberration and higher-order aberrations [13]. They also form a complete, orthogonal set, which means that any smooth function can be approximated up to an arbitrary precision given a large enough number of terms. We will use this property and rewrite the corrective phase profile as a sum of Zernike Polynomials:

$$\varphi(x,y) \cong 2\pi \sum_{i=4}^{N} \mathrm{z}_i \mathrm{Z}_i(x,y)$$

where Z_i is the ith Zernike Coefficient and Z_i is the ith Zernike Polynomial (using a single-numbering scheme described by Wyant and Creath [13] and used in ZEMAX as Zernike Fringe Phase [14]) and N is the index of the last term used in the expansion. We have ignored first three terms: piston and x- and y-tilt, hence, the summation starts at z_4 (focus coefficient).

B. Aberration-Corrected Hologram Generation

The image used as a reference is a single point in the replay field. The Fourier Transform of a single pixel is a grating, or a continuous phase surface if the phase can be unwrapped [1].

Therefore, the complex corrected hologram of the replay field can be rewritten as such:

$$\mathrm{H}(x,y) = \mathrm{e}^{i\Phi^{\mathrm{uv}}_{\mathrm{pixel}}(x,y)}\mathrm{e}^{-i\varphi}(x,y)$$

where H is the complex hologram, $\Phi^{\mathrm{uv}}_{\mathrm{pixel}}$ is the continuous phase surface corresponding to a pixel at a position (u,v):

$$\Phi^{\mathrm{uv}}_{\mathrm{pixel}}(x,y) = -2\pi\left(\frac{u}{u_{\max}}x + \frac{v}{v_{\max}}y + \varphi_{\mathrm{point}}\right)$$

and $\varphi(x,y)$ is the previously mentioned aberration-correcting phase mask composed of a weighted sum of Zernike Polynomials.

The complex hologram is then quantized due to the binary phase modulation of the SLM:

$$\mathrm{h}(x,y) = \begin{cases} 1, & \mathrm{real}(\mathrm{H}(x,y)) \geq 0 \\ -1, & \mathrm{real}\,(\mathrm{H}(x,y)) < 0. \end{cases}$$

To find the optimal correction hologram, we use a correction algorithm to find the values of z_i that minimize the fit function.

Once these are found, the mask can be straightforwardly applied thereafter to correct aberrations in every hologram generated for that particular projector:

$$H(x,y) = H_{\mathrm{uncorr}}(x,y)\mathrm{e}^{-i\varphi(x,y)} \tag{2}$$

where $H_{\mathrm{uncorr}}(x,y)$ is the arbitrary complex hologram to be corrected.

In order to prove that a hologram generated this way will correct the aberrations, one can apply (2) to (1):

$$\psi(u,v) = \mathcal{F}\left[B(x,y)H_{\mathrm{uncorr}}(x,y)e^{i\varphi(x,y)}\mathrm{e}^{-i\varphi(x,y)}\right].$$

It is noticed that phase aberrations with opposite signs cancel each other out inside the exponent, leaving the replay field as a Fourier Transform of the original, uncorrected hologram (as if the aberrations were not present):

$$\psi(u,v) = \mathcal{F}\left[B(x,y)H_{\mathrm{uncorr}}(x,y)\right].$$

III. Correction Procedure

As stated previously, we are trying to find a set of Zernike Coefficients for which the value of a fit function, representing the degree of residual aberration after correction, is the smallest. To

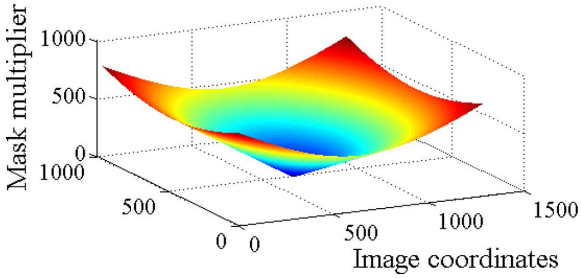


Fig. 2. Mask used to calculate the spread of pixels around the center. Points further from the center contribute more to the sum.

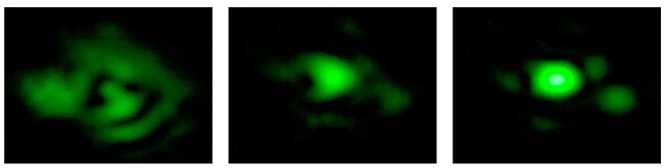


Fig. 3. Examples of points with different corrections. The respective fit function values of points are: (a) $= 4.7 \times 10^7$, (b) $= 1.5 \times 10^3$ and (c) $= 3.2$.

find a solution to this optimization problem, both a fit function and an appropriate optimization algorithm must be determined.

A. Fit Function

It is well known that an aberrated point will have lower peak intensity than the aberration-free point [12]. However, using a webcam with a bit-depth of 8, this information alone is insufficient for finding a perfect correction. Although this problem could be solved by using a sensitive photodiode, tilt aberration can shift the pattern and the results will become misleading. The tilt can also be corrected, but it would add another two other degrees of freedom.

In order to use a webcam for characterization of aberrations, a metric correlated with pixel's spread around the center is introduced

$$r = w_{\mathrm{blue}} \times \mathrm{w}_{\mathrm{green}} \times \sigma / q$$

where w_{blue} and w_{green} are weighting factors based on the point's intensity and σ is the weighting factor based on the spread of pixels around the center. When a point has high peak intensity and the pixels are circularly spread around the center within a small radius, the value of a fit function will be small. The normalization factor q is chosen for convenience.

Term w_{blue} arises from the realization that when the intensity of light is too high on the sensor to be recorded in the green channel, we begin to see high intensity in the other channels as well (blue has been chosen arbitrarily). The functional form of this weighting factor is

$$w_{\mathrm{blue}} = \begin{cases} 1, & \text{if } \max(\mathrm{I}_{\mathrm{blue}}) > 200 \\ 100, & \text{otherwise} \end{cases}$$

w_{green} in turn gives smaller values to the points for which the intensity in the green channel is higher. Its functional form is

$$w_{\mathrm{green}} = \left[256 - \max(I_{\mathrm{green}})\right].^2$$

When the highest value in the green channel is maximum (255), $\mathrm{w}_{\mathrm{green}}$ equals 1, otherwise it grows quadratically.

In order to calculate the spread of pixels around the center, we multiply the quantized image with the mask (Fig. 2)

$$\mathrm{mask}_{xy} = \sqrt{(x - x_{\mathrm{cent}})^2 + (y - y_{\mathrm{cent}})^2} + 1$$

where $(x_{\mathrm{cent}}, y_{\mathrm{cent}})$ are coordinates of the center of the pattern, which typically differ from the center of the image.

The pixel spread, σ is then equal to

$$\sigma = \sum_{\substack{i=0 \\ j=0}}^{\substack{i=x_{\max} \\ j=y_{\max}}} I_{\mathrm{quant}}(i,j) \times \mathrm{mask}(i + x_{\mathrm{ofs}}, j + y_{\mathrm{ofs}})$$

where $(x_{\mathrm{ofs}} y_{\mathrm{ofs}})$ is chosen so that the center of the mask overlays the center of the pattern and $I_{\mathrm{quant}}(i,j)$ is the quantized green channel of the image

$$I_{\mathrm{quant}}(i,j) = \begin{cases} 1, & \text{if } I_{\mathrm{quant}}(i,j) > I_{\mathrm{th}} \\ 0, & \text{otherwise.} \end{cases}$$

The threshold I_{th} can be adjusted so that good results are obtained. In our case, 20% of maximum intensity was chosen using trial and error method.

Equipped with this fit function we can examine the sample points, as shown in Fig. 3. It can be observed that there is a strong correlation between the quality of the point and the fit function value.

B. Hybrid Genetic-Steepest Descent Algorithm

The optimization method used here to find an optimal configuration of Zernike Coefficients is a hybrid between a Genetic Algorithm (GA) and Steepest Descent algorithm (SD). From initial experiments, GA was found to be very effective at finding new correction candidates, but was slow to optimize existing candidates. At that point SD was introduced to refine existing candidates to find the best corrections that lay nearby in the solution space.

Genetic mutations are an essential part of every genetic algorithm. Here, the role of mutations was replaced partially with SD optimization and partially by introducing purely random candidates into the population. The algorithm is illustrated in Fig. 4.

Each iteration of the algorithm produces 10 candidates, from which the best individual is chosen to use as the best correction found so far.

1) Genetic Optimization: The first step in the implemented genetic algorithm is non-deterministic tournament selection [15], [17]: among all the candidates a small number is randomly chosen (5–20, which corresponds to 0.2%–0.8% of the population, gives best results) and candidates are chosen with an appropriate probability. The one with the best value of the fit function is most likely to be chosen. The chosen candidate will become one of the parents. A similar procedure is used to choose the second parent.

Once appropriate parents have been selected, pairs of parents need to be 'mated': that is, have their parameters combined to form new potential offspring (solutions). In the majority of Genetic Algorithms, one or two children are produced from each pair of parents [11], [15], [16]. This is typically achieved by randomly selecting a single crossover point and swapping

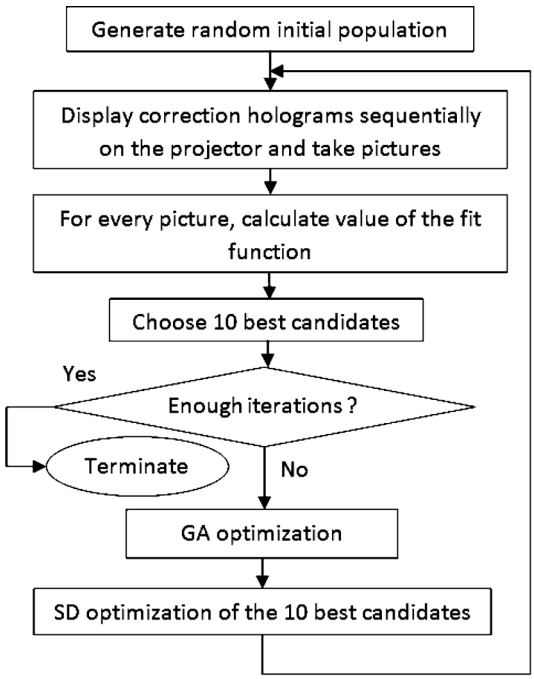


Fig. 4. Flow of a hybrid genetic-steepest descent algorithm.

TABLE I
NUMBER OF PERMUTATIONS DEPENDING ON THE TOTAL NUMBER OF VARIABLES

Number of variables 2n	Number of permutations $\binom{2n}{n}$
2	2
4	6
6	20
8	70
10	252
12	926

values from the parents after that point. However, with a single crossover at a random position, there is a very little chance that a resulting child would bring a positive improvement. Here, we present a different approach, intended to quickly generate a wider variety of children than standard approaches. In this method all the possible combinations of parents' parameters are generated producing children with on average half their parameters from each parent.

If we have 2n parameters, we must then choose n parameters to switch. The number of ways this can be done is the number of permutations: $\binom{2n}{n}$, shown in Table I.

It can clearly be seen that as the number of coefficients increases, the number of permutations also rapidly increases. We would like to incorporate as many parameters as possible. However, when we have more than 8 parameters, the numbers become impractically large and it would be too time consuming to test all the possibilities, taking into account that multiple crossovers need to be performed in a single iteration.

Therefore, Zernike coefficients have to somehow be put into 8 groups. The easiest way to do this is to identify the type of aberration they correspond to and group them accordingly as shown in Table II.

TABLE II
GROUPING ZERNIKE COEFFICIENTS z_4–z_{16} INTO 8 GROUPS

Group number	Aberration name	Zernike Coefficients
1	Defocus	z4
2	1st Order Astigmatism	z5, z6
3	1st Order Coma	z7, z8
4	1st Order Spherical	z9
5	1st Order Trefoil	z10, z11
6	2nd Order Astigmatism	z12, z13
7	2nd Order Coma	z14, z15
8	2nd Order Spherical	z16

```
for zIndex = 4 to 16
    for mult = -1 to 8
        set zOut to z
        set value to z(zIndex)
        if value<threshold
            set value to 0.125
        end
        set zOut(z) to value*mult/7
        generateHologram(zOut)
    end
end
```

Fig. 5. Pseudocode of the heuristic steepest descent algorithm.

The next step in the algorithm is the addition of random candidates into the solution population, which is used to introduce new genetic material into the system. This procedure has the same effect as introducing genetic mutations.

The number of random candidates should not be too large, otherwise the population will be purely random and this is not desirable. In this research the percentage of random candidates to the whole population was chosen to be 10% of the total population size.

It should be noted that the genetic algorithm presented here has been partially optimized for speed by generating a large number of holograms at once (and hence, minimizing the kernel's overhead). However, it is certain that other variants of Genetic Algorithms reported in literature [11], [14], [15] would also be suitable for the purpose of correction.

2) Heuristic Steepest Descent: Traditional steepest descent optimization relies on the ability to compute the derivative of the fit function. Here, the fit function is discontinuous at points where the light intensity changes. In order to avoid this problem, a novel heuristic steepest descent algorithm (HD) is proposed.

In the novel algorithm, the gradient is not calculated, but instead the algorithm tests all the principal directions in 13-dimensional space.

Let z(4) to z(16) be the Zernike Coefficients of the hologram to optimize and zOut(4) to zOut(16) coefficients of the output optimized hologram. Then the algorithm can be presented using pseudocode in Fig. 5.

Therefore, for each parameter: z(4) to z(16), the algorithm will test 9 possibilities: $-1/7, 0, 1/7, 2/7 \ldots, 1, 11/7$ of the original value, leaving the other coefficients unchanged. If a particular coefficient is smaller than some threshold, its value

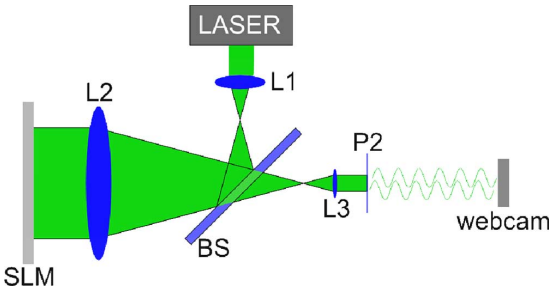


Fig. 6. Fourier projector facing a webcam (the same design as used by Freeman in [1]). The laser beam is expanded and collimated by lenses L1 and L2, reflected off a beamsplitter (BS), is modulated by the SLM in reflection mode, and is then demagnified by the combination of lenses L2 and L3. In plane P2 we see a reproduction of a hologram (with introduced phase aberrations), which undergoes diffraction and forms an image on the webcam.

TABLE III
TIMING OF PARTICULAR TASKS IN THE HYBRID GENETIC ALGORITHM

Task	Total time (s)	Time per hologram (ms)
Time of a single iteration	217	87.5
Hologram generation	13	5.2
Picture taking	166	67
Processing images	38	15.3

will be set to some defined number (0.125 was chosen as a result of a initial experiments).

IV. Experimental Setup

The setup consists of a Fourier projector and a webcam placed in front of it, shown in Fig. 6.

The main assumption made throughout this paper is that in the center of the replay field, the aberrations of the front lens (L3 in Fig. 6) are small and can be neglected. Therefore, all the aberrations will come from the non-flatness of the SLM and the imperfections in the other components of the setup, such as other lenses and mirrors.

A. Implementation and Timing

The hologram generation is implemented using nVidia's Compute Unified Device Architecture (CUDA) [18], resulting in very high speeds (5.2 ms per hologram). However, the majority of the time in each iteration of the algorithm is spent displaying holograms and taking pictures, as shown in Table III. This step could be further optimized using a faster webcam and by accessing the graphics memory directly.

B. Comparison of Correction Algorithms

In order to properly evaluate the performance of the hybrid genetic steepest descent and heuristic steepest descent algorithms presented here, it is necessary to compare their performance against conventional algorithm for a wide range of system aberrations. In reality, only two projectors with predetermined aberrations were available. In order to simulate a wider range of aberrations, many different random starting points (i.e., initial sets of Zernike Coefficients) are selected and the different algorithms are run using these initial values. Due to the fact that aberrations here are modeled as a linear superposition of Zernike polynomials, using a different starting point is equivalent to introducing an arbitrary optical aberration into the system and then attempting to correct it.

We used 15 random starting points to test the performance of the hybrid GA and heuristic SD in each case. The GA was allowed to run for 100 iterations, while HD was allowed up to 1000.

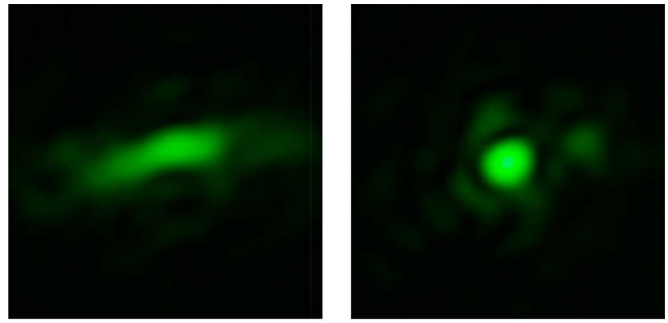


Fig. 7. Projector 1: uncorrected point (a) and a corrected one (b). The respective fit function values are: (a) $= 6.25 \times 10^7$, (b) $= 5.22$.

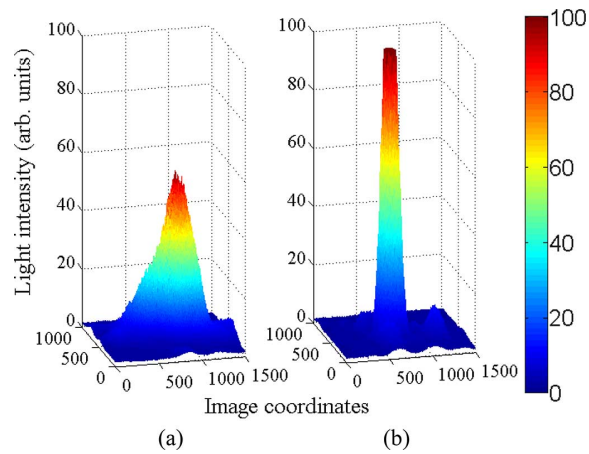


Fig. 8. Projector 1: uncorrected point (a) and a corrected one (b); surface view.

V. Results

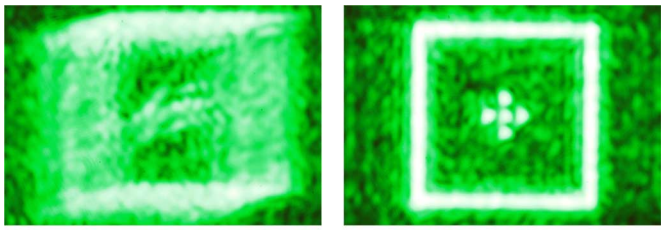
The algorithm is tested on two projectors and for three wavelengths. For the second projector, the result of the feedback loop is compared with the flatness of the SLM that was previously measured using an interferometric technique [1].

A. First Projector

The correction of this projector took 80 iterations $=$ 5 hours.

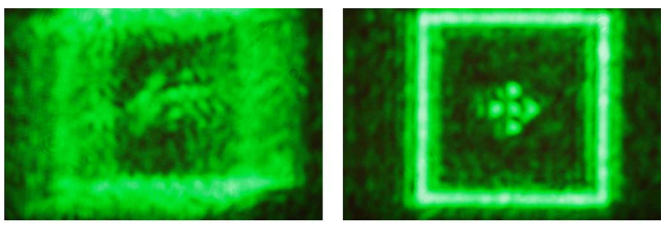
In Fig. 7(a) we can see the uncorrected image of a point. Point in Fig. 7(b) is the image of the same point with the applied correction and the improvement in quality is evident. Fig. 8 compares the intensity profiles of the two points and it is seen that the corrected point has high peak intensity and is quite narrow, whereas the uncorrected point has lower peak intensity (60%) and is smeared out.

As stated previously, the correction should be independent of the front lens. To test this hypothesis, we have taken pictures using two different lenses, one good quality lens and one cheap ball lens. The results can be seen in Figs. 9 and 10. A test image is a square with central pixels forming a cross in the middle



(a) (b)

Fig. 9. High quality CCD lens; square with a 5-pixel cross in the middle: (a) uncorrected versus (b) corrected.



(a) (b)

Fig. 10. Low-quality ball lens; (a) uncorrected versus (b) corrected.

(1 pixel apart). This way, we can test various properties of the image, such as horizontal lines, vertical lines, and interactions between single pixels.

It can be observed that the correction is similar for high-quality CCD lens (Fig. 9) as for the cheap ball lens (Fig. 10). In both cases the image is highly improved by correction. In particular it is visible that the central pixels are now separated from each other. The resultant corrective phase mask is shown in Fig. 11. It can be seen that one of the corners has a value close to 8. This is an error that results from the fact that we employ only 13 Zernike Polynomials to form a phase mask and higher order terms are necessary to control edge values. This error probably results in slight blurring of the image. However, less than 3% of the hologram is affected by this error, so its effect is likely to be small. This is verified by the quality of the corrected images [Figs. 9(b) and 10(b)], which do not exhibit substantial blurring.

It is also worth noticing that the SLM's irregularity, in classical terms, is not very significant. Excluding the erroneous corner, the peak-valley distance is of the order of one wavelength ($\sim$530 nm) and it results in complete blurring of the image to the extent that the shape of the image becomes unrecognizable.

B. Scaling of the Mask for Red and Blue Wavelengths

In order to project full-color images, as in real projectors, it is necessary to display three single-color images (red, green, and blue) by illuminating three holograms corrected using three phase-masks with appropriately colored lasers [6], [7]. This can be achieved by sequentially displaying the holograms on a single device and switching between illumination lasers. At a sufficiently high frame rate, the human eye will seamlessly merge these three single-color images together to form a full-color image. Ideally, it should be possible to adapt the aberration-correcting phase mask for these other wavelengths.

As stated in Section I, the phase change is proportional to the flatness difference as well as to the wavelength. Therefore,

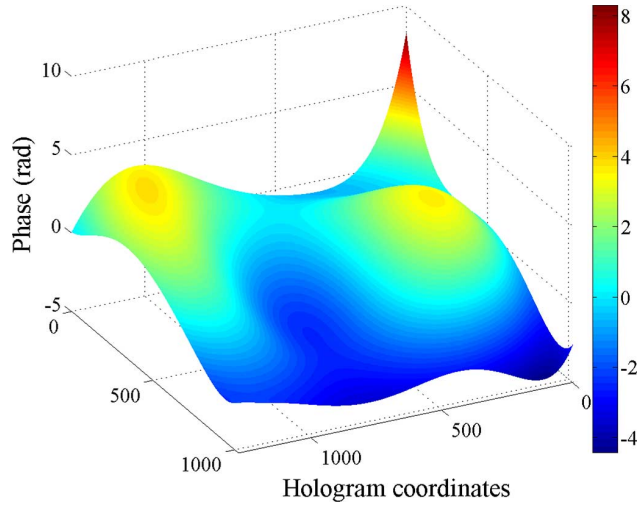
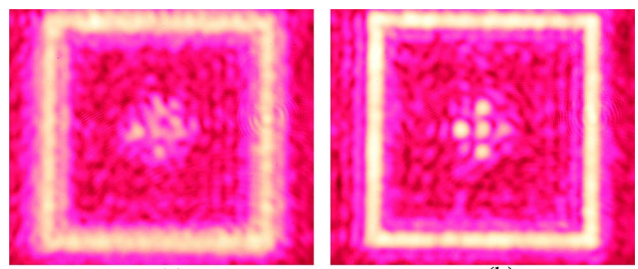
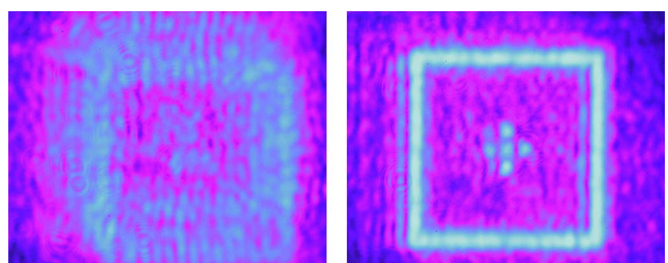


Fig. 11. Surface view of the phase mask for the first projector.



(a) (b)

Fig. 12. Red wavelength correction: (a) the same correction as green versus (b) rescaled mask.



(a) (b)

Fig. 13. Rescaled mask for blue light: (a) the rescaled and (b) re-focused (b).

if the correction was made for wavelength λ_1 and some other wavelength λ_2 is used to display the image, the mask has to be rescaled by the factor $\frac{\lambda_2}{\lambda_1}$. This procedure was successful in adapting the mask for red wavelength (Fig. 12).

With blue wavelength, due to a chromatic aberration of the lens, the image appeared out of focus [Fig. 13(a)]. This effect was easily eliminated by a simple re-focusing (adding the fourth Zernike Polynomial). After this procedure was carried out, the image was fully corrected [Fig. 13(b)].

C. Second Projector

The second projector tested had previously had the non-flatness of the SLM measured using interferometric techniques. Therefore, it is possible to use this as a reference to evaluate the performance of the new algorithm.

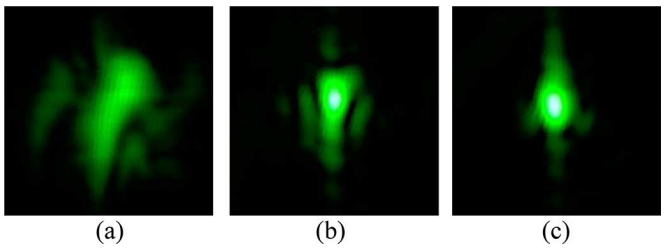


Fig. 14. Projector 2: (a) uncorrected point; (b) interferometric correction; and (c) feedback loop correction.

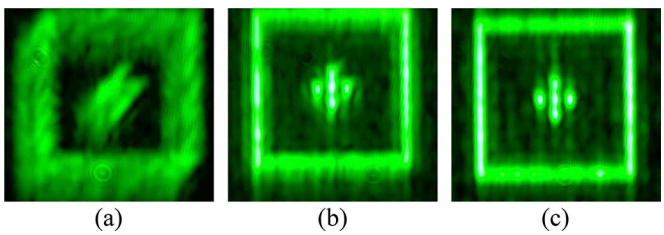


Fig. 15. Projector 2: (a) uncorrected shape; (b) interferometric correction; and (c) feedback loop correction.

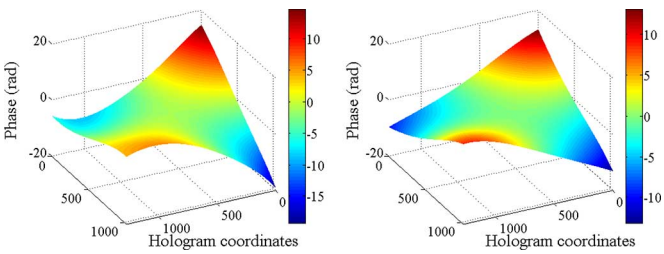


Fig. 16. Surface view of the phase mask: feedback loop mask (a) and interferometric mask (b).

The two flatness corrections are compared in Fig. 14 and it can be seen that feedback loop system [Fig. 14(c)] yields slightly better results than interferometric method [Fig. 14(b)] producing a more regular point with higher peak intensity.

Fig. 15 shows the correction of a square. Again, the feedback loop system yields a slightly better correction.

The corrective phase mask can be seen in Fig. 16(a) and can be compared with the mask determined from interferometric methods [Fig. 16(b)].

It is noted that the masks are broadly similar in shape. The feedback-loop system mask has higher curvature, which is due to the use of the limited number of Zernike Polynomials, but the superior results would suggest this is closer to the optimal correction. Additionally, the feedback-loop approach has the advantage of being fully automated.

VI. Comparison of Optimization Algorithms

The results for the fit function value versus time can be seen in Fig. 17 and the close-up in Fig. 18.

In most of the trials, convergence to the optimal correction is achieved. However, as seen in Fig.17, two trials of the GA test and one trial of HD test did not converge. This is likely the result of various internal and external factors. One potential external factor is the occasional failure of the laser used in the experiment. A likely internal factor is the domination of the

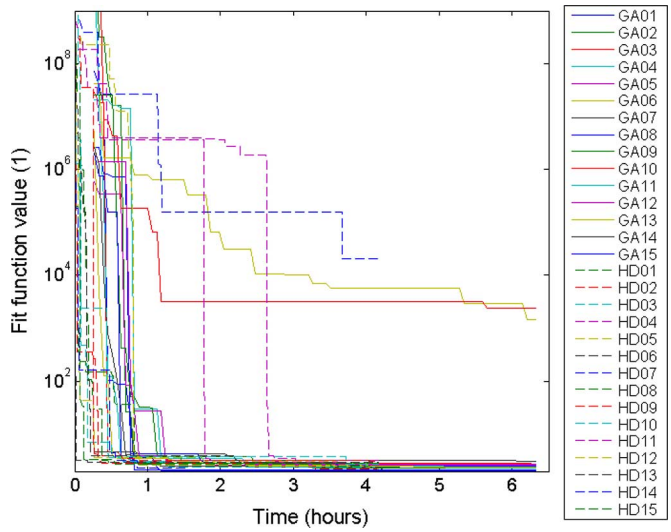


Fig. 17. Results of multiple corrections from Hybrid Genetic Algorithm (GA) and Heuristic Steepest Descent Algorithm (HD).

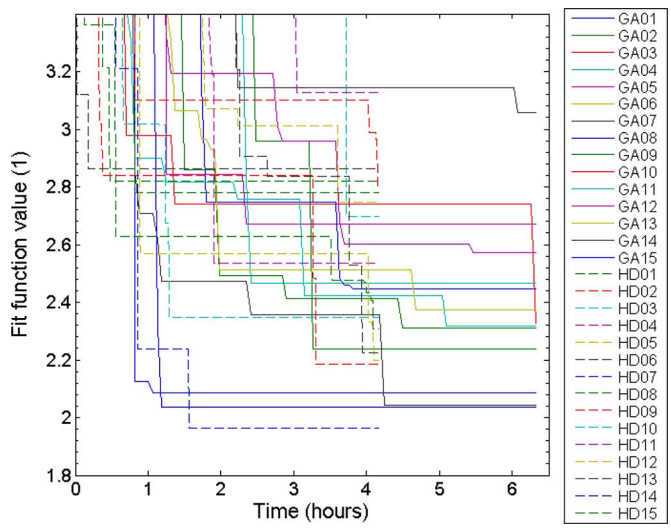


Fig. 18. Results of multiple corrections from Hybrid Genetic Algorithm (GA) and Heuristic Steepest Descent Algorithm (HD)—a close-up.

TABLE IV
Statistical Examination of Hybrid Genetic-Steepest Descent Algorithm and Modified Steepest Descent

Property	GA	MD
Minimum fit function value	2.04	1.96
Average fit function value	2.38	2.54
Standard deviation of fit function value	0.28	0.34
Average convergence time (hours)	3.8	2.6
Standard deviation of convergence time (hours)	1.7	1.5

population by similar solution candidates. This is partially addressed by introducing non-deterministic tournament selection in the GA and so, it is expected, that if left for longer, GA would eventually converge. However for HD it is rather more difficult to escape local minima.

The measured performance of the two proposed algorithms are summarized in Table IV. It can be seen that although the best GA correction is slightly worse than best HD correction, GA

performs slightly better than HD on average (smaller fit function mean and standard deviation). However, that comes at the expense of 50% longer convergence time (3.8 hours compared to 2.6 hours for HD).

Given that in reality the aberrations in a projector would not need to be corrected often, 4 hours is a reasonable time. Using state-of-the-art hardware and software, this time could be reduced to less than 1.5 hours.

VII. Conclusion

Two optimization-based methods of correcting aberrations coming from non-flat SLMs in Fourier projectors are presented. A specialized fit function, tailored to CCD sensors of limited bit depth, is used as the objective function that represents the degree of aberrations present. It takes into account two properties of aberrated single pixels: the peak intensity and the spread of pixels around the center. Using this fit function, a corrective phase mask that compensates aberrations is determined. Here, the mask is represented as a weighted sum of Zernike polynomials. These functions are well suited for this purpose, because they correspond to common optical aberrations. We approximate an aberration-correcting phase mask using 13 low-order Zernike Polynomials. The problem of aberration correction is then reduced to finding a point in 13-dimensional space of possible Zernike weightings that minimize the fit function. This is an optimization problem, with the aim of finding the mask producing the minimum value of fit function in the shortest period of time. To solve this optimization problem, a heuristic variant of the Steepest Descent algorithm, termed Heuristic Steepest Descent is introduced.

Another class of optimization algorithms that work well with similar types of problems are Genetic Algorithms. Here, we present a hybrid of a Genetic Algorithm (GA) and our Heuristic Steepest Descent Algorithm. GA is used in order to produce new correction candidates and HD is used to optimize existing candidates, thereby combining the most powerful features of both. The presented Hybrid GA algorithm is tested on two projectors and it is shown to produce high quality corrections after 5 hours for the first projector and 3.9 hours for the second. The corrective mask found by the algorithm for a given wavelength can be adapted for red and blue wavelengths by rescaling and compensating for defocus introduced by chromatic aberrations in the lens.

To test the performance of these two algorithms on different systems, they are both tested using several different starting points to mimic the effect of systems with different aberrations. The GA is found to perform better on average, with higher quality corrections produced, but at a cost of 50% slower convergence time.

References

[1] J. Freeman, "Visor projected helmet mounted display for fast jet aviators using a fourier video projector," Ph.D. dissertation, Dept. Elect. Eng., Cambridge Univ., Cambridge, U.K., 2009.

[2] T. D. Wilkinson, W. A. Crossland, and V. Kapsalis, "Binary phase-only 1/f joint transform correlator using a ferroelectric liquid-crystal spatial light modulator," *Opt. Eng.*, vol. 38, no. 2, pp. 357–360, Feb. 1999.

[3] N. J. New, "The binary phase only joint transform correlator," Ph.D. dissertation, Dept. Elect. Eng., Cambridge Univ., Cambridge, U.K., 2001.

[4] J. Carpenter and T. D. Wilkinson, "All optical mode-multiplexing using holography and multimode fiber couplers," *J. Lightw. Technol.*, vol. 30, no. 12, pp. –, Jun. 15, 2012.

[5] J. Carpenter, "Holographic mode division multiplexing in optical fibres," Ph.D. dissertation, Dept. Elect. Eng., Cambridge Univ., Cambridge, U.K., 2012.

[6] E. Buckley, "Holographic laser projection," *J. Display Technol.*, vol. 7, no. 3, pp. –, Mar. 2011.

[7] A. J. Cable, "Ph.D. Thesis," Ph.D. dissertation, Dept. Elect. Eng., Cambridge Univ., Cambridge, U.K., 2006.

[8] E. Buckley, "Computer generated holograms for real-time image display and sensor applications," Ph.D. dissertation, Dept. Elect. Eng., Cambridge Univ., Cambridge, U.K., 2006.

[9] C. S. Vikram, "Rayleigh versus Maréchal spherical aberration tolerance in in-line Fraunhofer holography," *Opt. Eng.*, vol. 33, no. 11, pp. 3715–3717, 1994.

[10] M. R. N. Avanaki *et al.*, "Sensor-less aberration correction in optical imaging systems using blind optimization," in *Proc. SPIE, 3rd Asia Pacific Opt. Sensors Conf.*, Jan. 31, 2012, vol. 8351, 83510B.

[11] P. Yang *et al.*, "A wavefront sensor-less adaptive optical system for a solid-state laser," *Opt. Laser Eng.*, vol. 46, pp. 517–521, 2008.

[12] J. W. Goodman, "Frequency analysis of optical imaging systems," in *Introduction to Fourier Optics*, 3rd ed. Englewood, CO, USA: Roberts, 2005, pp. 145–152.

[13] J. C. Wyant and K. Creath, "Basic wavefront aberration theory for optical metrology," *Appl. Opt. Opt. Eng.*, vol. XI, pp. 1–53, 1992.

[14] *Optical Design Program, User's Guide*. Redmond, WA: Radiant ZEMAX, May 16, 2002, pp. 135–138.

[15] K. V. Price, R. M. Storn, and J. A. Lampinen, "The differential evolution algorithm," in *Differential Evolution, A Practical Approach to Global Optimization*. New York, NY, USA: Springer, 2005.

[16] A. E. Eiben and J. E. Smith, "Genetic algorithms," in *Introduction to Evolutionary Computing*, 2nd ed. New York, NY, USA: Springer, 2007.

[17] B. L. Miller and D. E. Goldberg, "Genetic algorithms, tournament selection, and the effects of noise," *Complex Systems*, vol. 9, pp. 193–212, 1995.

[18] J. Sanders and E. Kandrot, "Parallel programming in CUDA C," in *CUDA by Example, An Introduction to General-Purpose GPU Programming*. Reading, MA, USA: Addison-Wesley, 2010.

A. Kaczorowski, photograph and biography not available at time of publication.

G. S. Gordon, photograph and biography not available at time of publication.

A. Palani, photograph and biography not available at time of publication.

S. Czerniawski, photograph and biography not available at time of publication.

T. D. Wilkinson, photograph and biography not available at time of publication.