A new Approach for Signal Loss Compensation in a Vibrometer*

Sascha Mayer, Bernd Tibken
Faculty of Electrical, Information and Media Engineering
University of Wuppertal
D-42119 Wuppertal, Germany
{mayer; tibken}@uni-wuppertal.de

Christian Rembe
Research & Development
Polytec GmbH
Polytec-Platz 1-7
D-76337 Waldbronn, Germany
C.Rembe@polytec.de

Abstract—This paper presents an approach for signal loss compensation in vibrometry, a contactless measurement technique. Since a new scanning method based on computer generated holograms was developed, now signal loss caused by speckle effects can be reduced and in the best case significantly compensated.

For this, multiple Zernike polynomials are used in a Zernike expansion mapped to a computer generated hologram, which modifies the wave front of a coherent laser beam. The coefficients of each Zernike polynomial are used as parameters in a global optimization. Solving this optimization problem with a suitable cost function leads to an increased signal level.

I. INTRODUCTION

To improve the quality of mechanical devices, investigation of vibration behavior is often useful. For instance, manufacturers of brake systems are highly interested in factors that cause squeaking noises. State-of-the-art methods to measure such vibrations are based on the Doppler Effect used in a Doppler-Laser-Vibrometer.

Thus, a laser beam is focused on the surface of a measurement object, the Vibrometer detects frequencies and amplitudes of mechanical oscillations [1], [2] by analyzing the shift in wave length of a reference beam that interferes with a reflected beam. Advantage of laser vibrometry is, that it is a contactless and non-reactive method.

Due to requirements of the customers, vibrometry needs scanning features which led to the development of Scanning-Vibrometers. Until now, Scanning-Vibrometers used galvanic mirrors [3] to place their measurement spots onto user-defined points on the measurement object. To scan the surface of the object with adequate speed, a position controller with small mechanical damping is necessary. These requirements induce additional signal noise in the detector signal which is caused by remaining torsional vibrations of the mechanical scanning system.

A novel technique for a Scanning-Vibrometer was presented in [3] to reduce these drawbacks. The mechanical scanning system was replaced by a Spatial Light Modulator (SLM). Computer Generated Holograms (CGH), written to this phase modulating SLM are utilized to modify angle, focus and position of a laser beam on a measured object.

Furthermore, with an SLM it is now possible to reduce the signal loss caused by the so-called speckle effect. Speckle patterns occur on rough grained surfaces, where the granularity size and the wave length of the laser beam are in the same range, which is discussed in [4].

The idea is to compensate such speckle patterns by writing correction patterns to the SLM based on the Zernike expansion [5]. Therefore, small aberrations will be mapped by Zernike polynomials which must be added to the current pattern on the SLM. Each Zernike polynomial is weighted by a Zernike coefficient that can be used as an optimization parameter. Thus, the reduction of signal loss of the detector signal leads to an optimization problem which can be solved by a standard optimization algorithm.

Section II presents a brief review of the mathematical concepts. Zernike polynomials and the DIRECT algorithm will be introduced. In section III the related optimization problem will be derived, section V summarizes the results of an experimental proof of concept which is presented in section IV. Finally, section VI gives a conclusion of the achieved results and an outlook to further research.

II. MATHEMATICAL CONCEPTS

A. Aberration patterns by using Zernike polynomials

In optical systems, aberrations can be described as a sum of weighted Zernike polynomials (see [5]) which represent wave front functions. These polynomials are a set of orthogonal radial polynomials, divided into odd and even polynomials and defined as

$$Z_n^m(\rho, \varphi) = \begin{cases} R_n^{m-2m}(\rho) \cos(m\varphi) & n - 2m \leq 0 \\ R_n^{m-2m}(\rho) \sin(m\varphi) & n - 2m > 0 \end{cases}, \quad (1)$$

where $R_n^{m-2m}$ is the radius-dependent part defined as

$$R_n^{m-2m} = \sum_{s=0}^{m} \frac{(-1)^s}{s!(m-s)!(n-m-s)!} \rho^{n-2s} \quad (2)$$

and $\rho$ is the normalized radial distance and $\varphi$ the azimuthal angle. Hence, the resulting Zernike expansion

$$\Phi(c, \rho, \varphi) = \sum_{n=0}^{k} \sum_{m=0}^{n} c_{mn} \cdot Z_n^m(\rho, \varphi) \quad (3)$$

is defined as sum of Zernike polynomials weighted with a related coefficient $c_{mn}$ for each polynomial. They can be

* This work was funded by the German Ministry of Education and Research (BMBF) under the grant number 13N9338.
collected in a vector \( c = (c_0, \ldots, c_{n_c}) \in \mathbb{R}^{n_c} \) where \( n_c \) is the number of used Zernike polynomials.

For further information on Zernike polynomials you may refer to [5]. Table I lists the Zernike polynomials used in our method.

### Table I

**OVERVIEW OF ZERNIKE POLYNOMIALS**

<table>
<thead>
<tr>
<th>#</th>
<th>( n )</th>
<th>( m )</th>
<th>name</th>
<th>polynomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>tilt Y</td>
<td>( \rho \cdot \sin (\varphi) )</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>tilt X</td>
<td>( \rho \cdot \cos (\varphi) )</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>astigmatism Y</td>
<td>( \rho^2 \cdot \sin (2\varphi) )</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>astigmatism X</td>
<td>( \rho^2 \cdot \cos (2\varphi) )</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>Coma Y</td>
<td>( (3\rho^3 - 2\rho) \cdot \sin (\varphi) )</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>Coma X</td>
<td>( (3\rho^3 - 2\rho) \cdot \cos (\varphi) )</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>2</td>
<td>spherical aberr.</td>
<td>( 6\rho^3 - 2\rho^2 - 1 )</td>
</tr>
</tbody>
</table>

The transformation of polynomials \( Z_n^m (\rho, \varphi) \) given in Table I from Polar to Cartesian coordinates with transformation rules from [6] leads to

\[
\Phi (c, x, y) = \sum_{n=0}^{k} \sum_{m=0}^{n} c_{mn} \cdot Z_n^m (x, y). \tag{4}
\]

The Zernike expansion in equation (4) represents a 2D phase aberration pattern which can be generated by CGHs on SLMs. Such 2D phase aberration patterns modify the incident wave front of a laser beam in its \( xy \)-plane. Therefore, the pattern is parameterized only by \( c \). With the assumption that the idealized laser beam has no shift in phase before the SLM, the solution of the homogenous wave equation with a constant phase shift \( \Phi \) following [7] is given by

\[
u(x, y, z, c, t) = e^{j(kz+\omega t+\Phi(c,x,y))}. \tag{5}\]

Since the phase offset behind the SLM is needed by a later introduced costfunction, we can define \( \Psi \) as constant phase pattern at \( t = 0 \) and \( z = 0 \) of the wave front in equation (5) as

\[
\Psi (c, x, y) = e^{j\Phi(c,x,y)}. \tag{6}\]

Itz is normalized to \(-1 \leq x \leq 1\) and \(-1 \leq y \leq 1\). Furthermore, an associated normalized aperture can be defined as

\[
\Gamma(x, y) = \begin{cases} 1 & -1 \leq x \leq 1, -1 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{7}\]

### B. DIRECT algorithm (one-dimensional)

With a given function \( f : D \subseteq \mathbb{R} \rightarrow \mathbb{R} \) an optimization algorithm has to solve the following problem in general

\[
f^* = \min_{x \in D} \{ f(x) \}. \tag{8}\]

In technical applications without closed mathematical descriptions, approaches for solving global optimization problems must be based on sampling methods, for instance the Monte-Carlo method [8] or the Shubert [9] algorithm. The Monte-Carlo method uses stochastic samples to find the global minimum. It is a non-deterministic blind-shot method. The maximum number of function evaluation per seconds in our technical system depends on the frame rate of the SLM. Currently, the used SLM can generate aberration patterns with a frequency of 30 Hz. So we can evaluate the function max. every 33 ms. Thus, we had to mind this time factor while choosing the optimization method. That is why a Lipschitzian method was selected.

The Shubert algorithm is a deterministic global optimization algorithm based on the Lipschitz constant \( K_L \) which is defined as

\[
K_L = \max \left\{ \frac{|f(x) - f(x')|}{|x-x'|} \right\}, \forall x, x' \in [x, \pi]. \tag{9}\]

All \( f(x) \) where \( \sup \{ f'(x) \} \) exists, which basically means all bounded, continuous functions fulfill this equation, and thus are called Lipschitz continuous.

With \( K_L \) defined in equation (9) as biggest absolute slope between any two points on the interval \([x, \pi]\) two simple linear bounding inequalities can be defined as

\[
f(x) \geq f(x) - K_L \cdot (x-x), \tag{10}\]

\[
f(x) \geq f(\pi) + K_L \cdot (x-x). \tag{11}\]

A lower bound \( y_{\min} \) for the minimum of \( f^* \) and its related \( x_{\min} \) are given by the point of intersection of the lines defined by equation (10) and (11) as follows

\[
x_{\min} = \frac{\pi + x}{2} + \frac{f(x) - f(\pi)}{2K_L}, \tag{12}\]

\[
y_{\min} = K_L \cdot \frac{x-\pi}{2} + \frac{f(\pi) + f(x)}{2}. \tag{13}\]

Figure 1 shows an example of the initial step of Shubert’s algorithm.

---

**Fig. 1.** initial step of Shubert’s algorithm
Unfortunately, the Shubert algorithm suffers from three major problems. Firstly, the objective function in our system is unknown and function values will be evaluated by measurements, so that the Lipschitz constant cannot be specified analytically. Secondly, for large Lipschitz constants the algorithm converges slowly and in n dimensions the algorithm needs $2^n$ samples to compute the next step.

The DIRECT algorithm overcomes these disadvantages. It was presented in [9] and modifies the Shubert algorithm.

Its idea is to shift the evaluation from the interval bounds to the center of the interval, given by $x_c = \frac{x + \pi}{2}$. Furthermore, equations (10) and (11) are adapted. This leads to new inequalities, given as

\[
\begin{align*}
  f(x) &\geq f(x_c) - K_L \cdot (x - x_c) \quad \text{for } x \in [x_c, \pi], \quad (14) \\
  f(x) &\geq f(x_c) + K_L \cdot (x - x_c) \quad \text{for } x \in [\pi, x_c]. \quad (15)
\end{align*}
\]

A lower bound for the minimum of $f(x) \forall x \in [\pi, \pi]$ is now given by

\[
y_{\text{min}} = f(x_c) - K_L \cdot \frac{\pi - x_c}{2}. \quad (16)
\]

In figure 2 changes towards to center point sampling has been made for the initial step.

![Fig. 2. initial step of the DIRECT algorithm](image)

With center point sampling, the method to split an interval $[x, \pi]$ has to change too. The DIRECT algorithm divides the interval $[x, \pi]$ into three parts of equal size such that

\[
[x, \pi] = \left[ x, \frac{2x + \pi}{3} \right] \cup \left[ \frac{2x + \pi}{3}, \frac{2\pi + x}{3} \right] \cup \left[ \frac{2\pi + x}{3}, \pi \right]. \quad (17)
\]

After dividing the current interval into three parts, an approach to select a good candidate for the next sampling is needed. We have to keep in mind that sampling is time consuming for our technical system and a good choice will speed up the algorithm. An adequate heuristic was also presented in [9].

Assuming that the start interval $[x, \pi]$ was trisected multiple times into m sub-intervals $[x_i, \pi]$, with their related center points $[x_{c_i}]$, for $i \in [1, m]$. Furthermore, $y_{\text{min}}$ is the current minimal sample. The sub-interval $[x_j, \pi]$, $j \in [1, m]$ is a good candidate if there exists a related slope $K_j > 0$ such that

\[
f\left([x_{c_i}]\right) - K_j \cdot \frac{x_{j} - x_c}{2} \leq f\left([x_{c_i}]\right) - K_j \cdot \frac{\pi_j - x_c}{2}, \quad (18)
\]

so interval $j$ has the smallest value $y_{\text{min}}$. Equation (18) describes candidates for next sampling by comparing the minimum (see equation (16)) of the considered interval $j$ with all other intervals. This requires that all positive slopes between such two points have to be analyzed. Figure 3 shows an example with some intervals, sorted with respect to the radius of it.

![Fig. 3. convex hull](image)

Candidates for next the sampling are those intervals which sample points form a partial convex hull on the set of points with positive slopes. Only the slopes between the filled dots satisfy the condition in equation (18). Thus, they are suitable slopes for a linear function that can be used to calculate a lower bound for all existing intervals.

To avoid a locally oriented search, a constant $\delta > 0$ was introduced. Equation (19) emphasizes samples that improves the lower bound by a significant value. Thus, $\delta$ is used to influence local vs. global search by setting percentually a threshold value that has to be exceeded.

\[
f\left([x_{c_i}]\right) - K_j \cdot \frac{x_{j} - x_c}{2} \leq y_{\text{min}} - \delta \cdot |y_{\text{min}}|, \quad (19)
\]

With the assumption that a list $\mathcal{L}$ can store samples in its form $\mathcal{S} = \{f(x_c), [x, \pi]\}$, the algorithm can be described as follows

1) **Initialize the DIRECT algorithm**

$\mathcal{L} = \{(f(x_c), [x, \pi])\}$ with $[x, \pi]$ the initial start box and $f(x_c)$ the measured sample at its center point.

2) **Identify all candidates for next sampling**

Test all candidates if they match the conditions in equation (18) and (19) for being good candidates, add them to a set $\mathcal{C}$ and remove them from $\mathcal{L}$.

3) **Split and sample**

For all candidates in $\mathcal{C}$:

a) Store list element as $\{f(x_c), [x, \pi]\}$ for temporary use and remove it from $\mathcal{C}$.

b) Divide interval into three parts, as shown in equation (17).

c) Evaluate sample of first and last new part at their center points. The middle interval has already evaluated and is given by the old $f(x_c)$ of the investigated interval.
d) Update $y_{\text{min}} = f(x_c)$ if $f(x_c) \leq y_{\text{min}}$.
e) Add new intervals to $\mathcal{L}$ including the middle interval.

4) Termination criterion
While the termination criterion (for instance: exceeding a max. number of iterations) is not reached, go to 2).

The multidimensional version of DIRECT can derived equivalently. Splitting domains leads to handle with hyper-rectangles and need further considerations that are described in [9].

III. RELATED OPTIMIZATION PROBLEM

A. Associated system

Equation (6) represents the normalized square of a 2D phase aberration pattern that can be varied only by the related coefficients $c \in \mathbb{R}^n$. These Zernike expansion coefficients are used to modify the CGH on the SLM and are the independent variables of the system we want to optimize. At first, we have to take a look at the associated system in figure 4.

![System diagram](image)

Fig. 4. System to be optimized

The measured system output is the signal strength on a logarithmic scale. Variation of $c$ changes the signal strength but, unfortunately, also the position and shape of the laser spot. To avoid large variances in position and shape, a cost function consisting of the detected signal and measure of spot quality is needed. Let $\alpha(c)$ be the measured signal and $\beta(c)$ a measure for spot quality, then we can assume for a first approach the following cost function

$$z(c) = \gamma \cdot \frac{1}{\alpha(c)} + \frac{1}{\beta(c)}.$$ (20)

On the one hand, $\beta(c)$ could be calculated from the aberration pattern on the SLM; on the other hand $\alpha(c)$ cannot be identified analytically because it depends on the surface of the measurement object and its microscopic structure. Accordingly, a derivative of the cost function cannot be calculated which led to the choice of the DIRECT algorithm.

B. Measure of spot quality

The quality measure $\beta(c)$ should describe the intensity level within the aperture with emphasis onto the center. From [10] it is well known that the intensity distribution of the light within an optical path limiting aperture $\Gamma$ is proportional to the 2D Fourier transform of itself, given by

$$\Gamma_F(u, v) = \mathcal{F}\{\Gamma(x, y)\}.$$ (21)

For use in a measure of quality, reduction to a real number is necessary. Therefore the square of $\Gamma_F(u, v)$ will be used.

The following equation defines $\beta(c)$ by weighting $\Gamma_F(u, v)$ with the Fourier transform of $\Psi(c, x, y)$

$$\beta(c) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Psi_F(c, u, v)|^2 \cdot |\Gamma_F(u, v)|^2 \frac{dudv}{\max\{|\Gamma_F(u, v)|^2\}}.$$ (22)

normalized to the maximum value of $|\Gamma(u, v)|^2$. $\Psi_F$ can also be written as

$$\Psi_F(c, u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\Psi(c, x, y) - 2\pi(ux + vy))} dx dy.$$ (23)

For Zernike polynomials of order 3 and above this integral cannot be solved analytically, thus a discrete Fast Fourier Transform (FFT) must be performed in the algorithm. However, equation (21) leads to a closed analytical expression under the assumption that the aperture is limited by a normalized square

$$\Gamma_F(u, v) = \int_{-1}^{1} \int_{-1}^{1} e^{-j2\pi(ux + vy)} dx dy$$

$$\Rightarrow \Gamma_F(u, v) = \frac{1}{\pi^2 uv} \cdot \sin(2\pi u) \cdot \sin(2\pi v).$$ (24)

Figure 5 shows $\Gamma_F(u, v)$. It is easy to see that the square of the intensity distribution $|\Gamma_F(u, v)|^2$ in equation (22) will emphasize the values near the center of the aperture similarly to a 2D Gaussian filter.

In our technical system, $\beta(c)$ will be calculated from the current phase aberration pattern $\Psi(c, x, y)$ written to the SLM and the discrete representation of the analytical expression of equation (24).

![Gamma_F(u,v) plot](image)

Fig. 5. plot of $\Gamma_F(u, v)$

C. Optimization problem

The given optimization problem belongs to the class of multi-objective optimization. Therefore, in equation (20) $\gamma \in \mathbb{R}$ is a constant scalar to shift between the objectives. Finally, the optimization problem can be written as follows

$$c^* = \arg\min_c \left\{ \gamma \cdot \frac{1}{\alpha(c)} + \frac{1}{\beta(c)} \right\}.$$ (25)
The used Vibrometer with applied SLM has been developed by Polytec with collaboration of the Institute of Technical Optics (ITO) of the University of Stuttgart, Germany. It is based on a Polytec IVS 300 [3].

Initial measurements have been obtained under laboratory conditions. A half-transparent plastic plate in front of a mirror was optimal for first experimental investigations and led to a stable damped signal. Figure 6 shows a sketch of the hardware setup.

![Holo-Vibrometer](image)

**Fig. 6.** Hardware setup

This hardware was set up on a special vibration absorbing table. First tests with a mirror have shown that there is only a minimal speckle effect caused by the surface of a mirror. This leads to the conclusion that using a mirror in the hardware setup did not influence the following results.

The used SLM is limited in its frame rate which is currently 30 Hz. At first, an investigation of the time delay of the system has to be done so that the evaluation rate of the algorithm did not violate this latency.

Therefore, the laser spot was switched between two spot sites. The first spot site generated a strong signal while the second site generated a weak one. The spot was switched every 500 ms between these sites and a measurement of the signal level was done every 5 ms. Figure 7 shows the latency analysis for 5 steps.

![Latency analysis](image)

**Fig. 7.** Analysis of the system latency (5 steps)

The average system latency was calculated by averaging over 100 steps and led to system latency of 86 ms (see figure 8). This latency has to take into account while executing the optimization algorithm. That means that after writing a new pattern to the SLM, the algorithm has to wait at least 86 ms before it measures the signal strength.

For this hardware setup a closed-loop system was developed which applies the optimization algorithm. Controlling will be stopped by reaching a given threshold or maximum number of feedback steps. An extremum seeking control is not feasible at the moment because the frame rate of the SLM is too slow with respect to the scanning frequencies. A sketch of the deployed system can be seen in figure 9.

![Closed-loop system](image)

**Fig. 9.** Related closed-loop system

To obtain reasonable optimization results, the parameter range of the Zernike coefficients had to be constrained. These constraints were identified by a special test series. With this preparations the optimizer was tested by several measurement experiments.

A software written by ITO was used to control the SLM. It allows setting and modifying aberrations to the SLM. To calculate the necessary FFTs, the library for Fast Fourier Transformation fftw3 [11] was used. Furthermore, a standard optimizer for nonlinear programming was used. It is available with the well known nlopt library [12]. Both libraries are published under the GNU GPL and are freely available for educational and commercial use.

The detector board provides the heterodyne carrier signal for the FM (frequency modulation) demodulator and a voltage signal which is a measurand for the signal strength of the carrier. The signal strength corresponds to the detected light power. It delivers values in the range of 0 to 5 V. The voltage signal is proportional to the signal strength with a factor of 6 dB/V in the linear range. Thus, an increase of 500 mV corresponds to an increase of the SNR (signal-to-noise ratio) by a factor 2. The data acquisition was conducted with a National Instruments Ni DAQ-Pad.
The software was developed with the integrated development environment Visual Studio 2008.

V. OPTIMIZATION RESULTS

For investigation of convergence behaviour and effectiveness of the algorithm, 4 passes with different polynomials were performed. For the first optimization pass Zernike polynomials 1 to 4 as described in table I were used. The second optimization pass used polynomials 1, 2, 5, 6. Both passes were weighted with $\gamma = 2$, aborted after 100 iterations and led to an improvement of the signal strength. Whereas the first pass increased the detector signal by about 1.27V which means an increase of the SNR by a factor of 5.78. Pass 2 could increase the detector signal by only about 0.92V (increase of the SNR by a factor of 3.35).

Both passes led to an increase of the detector signal by about 1.0V (increase of the SNR by a factor of 4).

Pass 3 reached the optimized signal strength after 300 iterations while pass 4 needed about 370 iterations. Figure 11 shows both passes, where the solid graph represents the third and the dashed graph the fourth pass.

VI. CONCLUSION AND OUTLOOK

It has been shown that modifying the wave front of a laser beam by mapping aberration patterns to an SLM can improve the signal level significantly. The choice of the polynomials and their number is important for the convergence of the approach. The best increase of the signal level was reached by pass 1. Pass 3 and 4 also led to an increased signal level but the laser spot was massively deformed.

Additionally, the DIRECT algorithm is a suitable candidate for global optimization for this problem. Next idea in research is to adapt the algorithm to the given problem. Therefore, the DIRECT algorithm has to be improved in 3 respects:

- Weight of optimization parameters must be possible.
- Modification of DIRECT for better convergence.
- With respect to multi point methods for vibrometry, DIRECT has to be analyzed if parallel processing (multiple instances on CUDA) is possible.

Also, measurements will be extended to find the optimal polynomials with the best convergence.

Another issue of research could be the investigation of filters that replace the intensity distribution in equation (22) and which could be applied in time domain, so a FFT would not be necessary. Furthermore, methods to reduce the noise on the current measured sample should be investigated, for instance with an adapted Kalman filter.

REFERENCES


