PROCEEDINGS OF SPIE

Increase of wafer inspection tool throughput with computational imaging

Oleg Soloviev, Hieu Thao Nguyen, Jacques Noom, Michel Verhaegen

SPIE.

Increase of wafer inspection tool throughput with computational imaging

Oleg Soloviev[a,b], Nguyen Hieu Thao[a,c], Jacques Noom[a], and Michel Verhaegen[a]

[a]DCSC, TU Delft,Mekelweg 2, 2624 CD Delft, the Netherlands
[b]Flexible Optical BV, Polakweg 10–11, 2288 GG Rijswijk, the Netherlands
[c]RMIT University Vietnam, 702 Nguyen Van Linh St, District 7, Ho Chi Minh 70000, Vietnam

ABSTRACT

To increase the throughput of image-based wafer quality inspection tools, we propose to use computational imaging methods to address two main bottlenecks: the mechanical alignment of the wafer with the imaging plane and pixel size of the imager. The former requires significant time but is crucial for a reliable quality check, and the latter introduces a trade off between the wafer scanning speed and the minimum detectable defect size.

We demonstrate application of our recently developed SANDR algorithm for obtaining a wafer image with sub-pixel resolution from a series of not-perfectly aligned low-resolution images. The wafer misalignment creates a varying over the field of view depth-of-focus, which poses a significant obstacle for the state-of-the-art methods, but is successfully processed with SANDR. The method is tested on simulated images.

Keywords: Deconvolution, inverse problems, wafer inspection, sub-pixel resolution, non-uniform blur

1. INTRODUCTION

Image-based quality control is one of the important tools used during the manufacturing and end quality checks in semiconductor,[1] automotive,[2] and many other industries. For Industry 4.0, requiring fully automatized quality checks, the spatial resolution (size of the smallest feature that can be inspected) is the key factor that affects the overall efficiency and throughput of the control tool. As an example, in semiconductor industry, where the feature size gets smaller and smaller, the optical inspection ('macro inspection') becomes both more challenging and crucial for defect-free and high-yield wafer manufacturing.[3]

For high-end imaging systems used in these tools, the spatial resolution is proportional to the quotient between the pixel size of the camera and the optical magnification, and thus for a higher resolution, either a smaller pixel size or a larger magnification is required. Larger magnification corresponds to smaller field of view, which in turn requires more scanning positions and leads to the increased inspection time and decreased throughput. Therefore, the preference is often to have a smaller pixel size. However, there are technological limits to the magnification and the smallest pixels that can be manufactured and/or used in these tools. Hence a computational approach to increasing spatial resolution provides an interesting alternative.

Under certain conditions, a high resolution image of an object can be restored from one or several low-resolution (LR) images. In particular, if sub-pixel image displacement is present in a series of LR images, a synthetic fused image sampled on a denser grid can be computed by placing the pixel values for the coarsely sampled images in the appropriate places and filling in the gaps using the available *a priori* knowledge.[4] This reconstruction of an object with high-resolution from several low-resolution images capturing the object at subpixel-offset positions, called the (pixel) Super-Resolution (SR) problem, has been studied for many years,[5,6] and a number of solution approaches have been proposed for the SR problem, including direct methods[7,8] and iterative algorithms.[9–11] According to a recent overview [4, Chapter 4], both approaches can be classified as a special case of underdetermined inverse imaging problems, the arbitrary sampling and bounded spectrum reconstruction (ASBSR).

Further author information: (Send correspondence to O.S.)
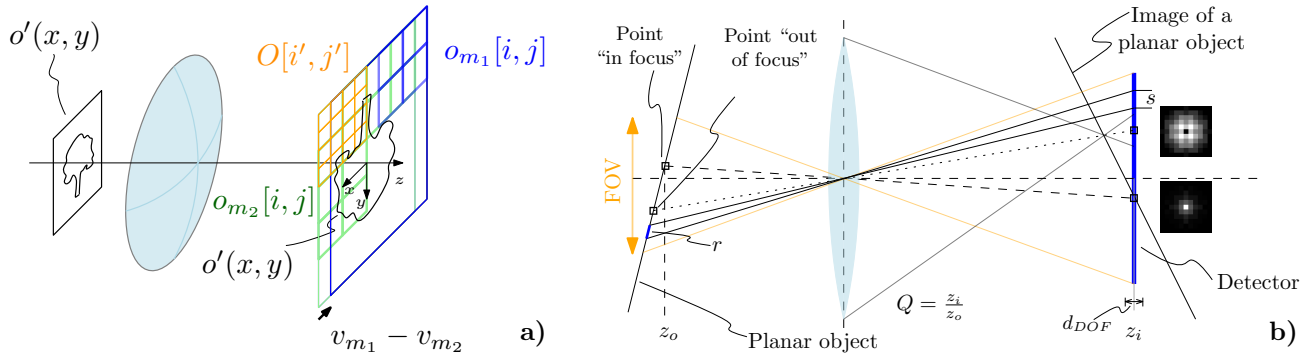O.S.: E-mail: o.a.soloviev@tudelft.nl

Figure 1. a) Imaging scheme used in the problem formulation. An imaging lens with magnification Q creates image $o'(x, y) = o(x/Q, y/Q)$ of some planar object $o(x)$, which is registered by a detector with pixel size s. $O[i', j']$ denotes sampled with rate s/τ values of $o'(x, y)$ (the fine grid is shown in orange). To obtain O, M sampled images $o_m[i, j]$ (green, blue), with introduced subpixel offsets v_m in each of them are recorded. The superresolution problem is to restore $O[i', j']$ from $\{o_m\}$. b) Side view of the imaging scheme with an example of misalignment of the object and detector planes creating position-dependent defocus blur.

SR reconstruction is possible if the LR images are registered for different subpixel-offset positions of the object. In practice, shifting the object at subpixel scale can be a major challenge to the SR problem and gives rise to several important questions that need to be addressed. Camera shake and motion blur induced by the shifts and inaccuracy of subpixel registration has been considered in the literature. Ideally, the shifting process should not cause any variations in the object orientation with respect to the camera.

However, this is not always achievable in practice, especially in applications with large magnifications, and correspondingly large values of the numerical aperture and shallow depth of focus on the detector. While introducing the subpixel shifts, the misalignment of the object with the best imaging plane can occur. This can introduce a (non-uniform) defocus blur in each of the LR images, which will affect the SR image restoration. To the best of our knowledge, the latter challenge has not been considered in the literature of SR by subpixel motions.

In this manuscript, we present our current progress in the task and demonstrate the first results of the computational super-resolution obtained from four LR images with sub-pixel shifts and non-uniform defocus.

2. SUPER-RESOLUTION WITH SUB-PIXEL SHIFTS

In this section we describe our approach and formulate it as an inverse problem.

2.1 Outline of the method

The imaging model with introduced subpixel shift is shown in Fig. 1. Let $o'(x, y) = o(x/Q, y/Q)$ be an optical image of some object $o(x, y)$ created with magnification Q and let s be the pixel size of the detector. To increase the field of view in τ times, where τ is some integer, the optical magnification should be decreased τ times. The (pixel) resolution is defined by the minimum feature size resolved by the sampling grid (here in all cases we consider the diffraction-limited PSF to be unresolved by the pixel size), consequently to keep the minimal resolved feature size, the sampling rate should be also increased τ times. The (pixel) super-resolution is the problem of obtaining "high-resolution" (that is sampled on a fine grid with step s/τ) image from one or several "low-resolution" images (sampled on the detector grid with pixel size s).

An optical image of an object $o(x, y)$ is sampled with a detector with a pixel size s able to create an image $o[i, j]$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, which we will denote as "low-resolution" one. To obtain a "high-resolution" sampling $O[i', j']$, $i' = 1, \ldots, \tau I$, $j' = 1, \ldots, \tau J$ on a finer grid with sampling rate s/τ, where τ is some positive integer, several low-resolution images $o_m[i, j]$, $m = 1, \ldots, M$ are recorded. Before taking every image, a shift of the imaging detector v_m is introduced (while shifts of arbitrary nature can be used,[4] here for simplicity and to

limit the number of the images, we consider fixed values of the shifts). In this work, we consider $M = \tau^2$ shifts with values from the finer grid

$$v_m \in \left\{ \frac{s}{\tau}(n_x, n_y) \mid n_x, n_y = 0, \ldots, \tau - 1 \right\}.$$

Let us define the downsampling operator with rate τ, D_τ, through binning operation on tiles of size $\tau \times \tau$, that is

$$D_\tau : \mathbb{R}^{\tau I \times \tau J} \mapsto \mathbb{R}^{I \times J} \tag{1}$$

for any $I, J \in \mathbb{N}$ and if $a = D_\tau(A)$, then

$$a[i,j] = \sum_{n_x, n_y = 1, \ldots, \tau} A[(i-1)\tau + n_x, (j-1)\tau + n_y]. \tag{2}$$

Then in the noiseless case each of the low-resolution images can be represented as downsampling of a shifted high-resolution image

$$o_m = D_\tau \circ T_m(O), \ (m = 1, 2, \ldots, M), \tag{3}$$

where $T_m = T_{(n_x, n_y)}$ is a zero-padded shift operator, that is if $A = T_{(n_x, n_y)}(B)$, then

$$A[i,j] = \begin{cases} B[i - n_x, j - n_y], & i - n_x \geq 1, j - n_y \geq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For unknown values of $O[i', j']$ Eq. (3) represents a square linear system of equations with $\tau^2 IJ$ unknowns, which can be solved using least-squares method, also for noisy measurements of o_m.[5,12]

2.2 Hardware

The subpixel shifts can be introduced by several methods:

1. moving the object under inspection,

2. moving the detector,

3. moving the optical image on the detector.

The last option can be in turn, implemented using two folding mirrors mounted on translation stages, each of which introduces shift in x or y direction or with a Porro prism of the second type,[13] as shown in Fig. 2.
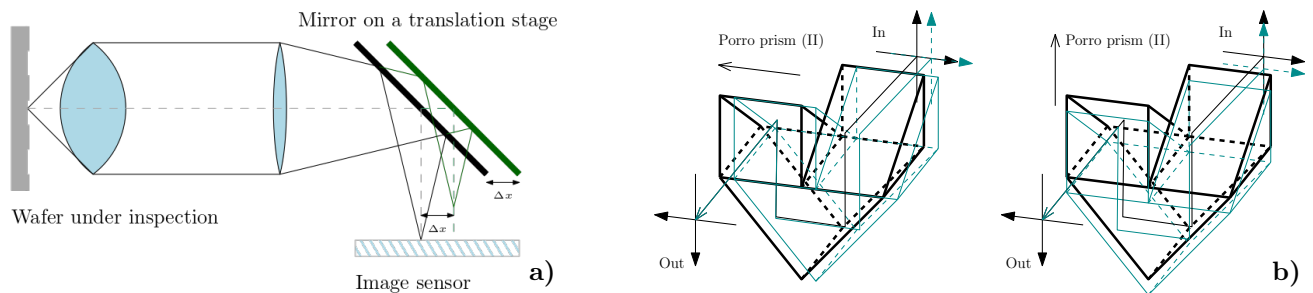

Figure 2. Introducing a) unidirectional and b) bidirectional image shifts with folding mirror and Porro prism of the second type. Image b) is redrawn after Ref. 13.

A particular implementation depends on a number of factors which are out of the scope of this manuscript.

2.3 Forward model for the images

Depending on a particular implementation, uniform or non-uniform defocus might be introduced as a result of image subpixel shifts. For our use case of wafer inspection, the most general model would be to represent the out-of-focus blur as changing in one direction across the image, see Fig. 1 (b). In this case the local PSF $p = p(d)$ is defined by the local defocus value $d(x, y)$ constant along some direction:

$$p(x, y; d) = \left| \mathcal{F} \left(A \mathrm{e}^{\mathrm{i} d Z_2^0} \right) \right|^2, \quad d = d(\mathbf{b} \cdot (x, y)), \tag{5}$$

where Z_2^0 is the Zernike polynomial of order two and azimuthal frequency zero (defocus), and $\mathbf{b}$ is some vector defined by the misalignment. Then the optical image $i(x, y)$ in the detection plane is represented by anisoplantic convolution

$$i(x, y) = \int o'(x', y') p(t_x - x, t_y - y; \mathbf{b} \cdot (x, y)) \, \mathrm{d}t_x \, \mathrm{d}t_y. \tag{6}$$

In practice, we can simplify the equation above by discretization of values of d, for instance, considering

$$d_n = n d_{DOF}, \tag{7}$$

where d_{DOF} is the defined by the depth of focus of the imaging system (d_{DOF} can be also an arbitrary value which is "tolerated" by the following decision-making system or a human operator). By defining the zones of equal defocus Δ_n as

$$\Delta_n = \{[i, j] \mid (n - \frac{1}{2}) d_{DOF} \leq d(s \cdot [i, j]) < (n + \frac{1}{2}) d_{DOF}\}, \tag{8}$$

we can subdivide each image i_m in the zones of equal defocus μ_{mn} defined as

$$\mu_{mn}[i, j] = \begin{cases} 1 & \text{if } (i, j) \in \Delta_n, \\ 0 & \text{otherwise}, \end{cases}$$

and we consider the blur to be constant inside each of the zone and denote it as p_n.

Then for each image i_m the integral of (6) can be approximated as

$$i_m = B_m(o_m) = \sum_{n=1}^{N} (\mu_{mn} \cdot o_m) * p_n, \ (m = 1, 2, \dots, M), \tag{9}$$

where we have introduced operator B_m to represent the effect of anisoplantic blur.

In the literature devoted to non-uniform DOF blur removal, this approach is often used to separate one or several "foreground" objects from the "background". Then in each of the zones of equal defocus, the blur can be removed by isoplanatic methods and the estimated patches of the scene are stitched together. An important difference of our approach is that for a planar object and a shallow depth of focus, zones μ_{mn} are represented by quite narrow bands in the image, with width comparable with the blur width, and the standard approaches based on stitching of isoplanatically deconvolved areas are not applicable any more. A novel approach to address this issue was developed by us recently.[14]

Finally, combining Eqs. (3) and (9), we have the following forward model connecting the recorded with subpixel shifts LR images i_m and the HR image O:

$$i_m = B_m \circ D_\tau \circ T_m(O). \tag{10}$$

3. SOLVING THE INVERSE PROBLEM SANDR ALGORITHM

The goal of this work is to solve the inverse to (10) problem of reconstructing a SR Image using Non-uniform Defocused (SIND) images, called the SIND problem.

The SIND problem can be subdivided in three sub-problems: 1) estimation of non-uniform defocus models, 2) removal of non-uniform defocus effects, and 3) reconstruction of an SR image. In this work, we address the last two tasks and propose a new algorithm to simultaneously perform both Superresolution reconstruction And Nonuniform Defocus Removal (SANDR).

The SANDR algorithm combines our recent nonuniform defocus removal method[14] and the least squares approach[15] for subpixel image reconstruction. The key feature of the proposed algorithm is that the methods are combined not in a sequential manner. Hence, it inherits global convergence from its two component techniques and avoids the typical error amplification of multi-step optimization contributing to its robustness. Further, existing acceleration techniques for optimization[16] have been used that assure faster convergence of the SANDR algorithm compared to most existing SR techniques.

In assumption of i.i.d. noise present in i_m measurements in Eq.(10), the maximum-likelihood approach leads to the following minimization problem:

$$\min_{O \in \Omega} \ f(O),$$ (11)

where f represents the data fidelity given by

$$f(O) = \sum_{m=1}^{M} \|i_m - B_m \circ D_\tau \circ T_m(O)\|^2,$$ (12)

and Ω is a set of all feasible objects (that is all real-valued matrices with elements in segment $[0,1]$). Using this definition, the constrained optimization problem of Eq.(11) can be written as unconstrained optimization

$$\min_O \ f(O) + \iota_\Omega(O),$$ (13)

where $\iota_\Omega(x)$ is the indicator function of set Ω:

$$\iota_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$ (14)

Note that in this form the optimization problem is formulated simultaneously for the problems of Superresolution reconstruction And Non-uniform Defocus Removal (SANDR), as opposed to a possible approach of firstly de-blurring the LR images and then reconstructing the SR image from them (in the next section, we will refer to this approach as "Sequential Minimization"). In addition, Eq. (13) can be seen as minimization of the data fidelity term and regularization term, for which a number of techniques, *e.g.* proximal methods,[17] exists.

In this work, we have applied the FISTA[16] method. Making use of the acceleration techniques for optimization of FISTA assures fast convergence of the SANDR algorithm going from rate $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$ compared to most existing SR techniques using the gradient descent method.

3.1 Algorithm description

For simplicity, in the description below we made some abuse of notation by denoting through T_{-m} an inverse operator to T_m, that is if $T_m = T_{n_x,n_y}$, then $T_{-m} = T_{-(n_x,n_y)}$. By U_τ we denote an operator that gives the least-squares solution to the equation $y = D_\tau(x)$, namely if $A = U_\tau(a))$, then

$$A[(i-1)\tau + n_x, \ (j-1)\tau + n_y] = \frac{a[i,j]}{\tau^2} \quad \forall n_x, n_y = 1, \ldots, \tau.$$ (15)

Algorithm 1 (SANDR).
Input: i_m – LR images, B_m – blur operators, λ – stepsize, $t^{(0)}$ – initial acceleration stepsize, K – number of

iterations, and $\varepsilon > 0$.

Initialization: $X^{(0)} = O^{(0)} = \frac{1}{M} \sum_{m=1}^{M} T_{-m} \circ U_\tau(i_m)$.

Iteration process: given $X^{(k)}, O^{(k)}, t^{(k)}$

$$G_m^{(k)} = T_{-m} \circ U_\tau \left(\nabla f_m \left(D_\tau \circ T_m \left(O^{(k)} \right) \right) \right),$$

$$X_m^{(k+1)} = P_\Omega \left(O^{(k)} - \lambda G_m^{(k)} \right),$$

$$X^{(k+1)} = \frac{1}{M} \sum_{m=1}^{M} X_m^{(k+1)},$$

$$t^{(k+1)} = \frac{1 + \sqrt{1 + 4t^{(k)^2}}}{2},$$

$$O^{(k+1)} = X^{(k+1)} + \frac{t^{(k)} - 1}{t^{(k+1)}} \left(X^{(k+1)} - X^{(k)} \right).$$

Stopping criteria: $k > K$ or

$$\sum_{m=1}^{M} \left\| G_m^{(k)} \right\| < \sum_{m=1}^{M} \left\| G_m^{(k-1)} \right\| + \varepsilon. \tag{16}$$

Output: $\widehat{O} = P_\Omega \left(O^{(\texttt{end})} \right)$.

Above, P_Ω is the projection operator on set of all feasible objects Ω and the functions f_m are given by

$$f_m(x) = \frac{1}{2} \left\| B_m(x) - i_m \right\|^2, \ (m = 1, 2, \dots, M).$$

For the details on efficient evaluation of ∇f_m which uses the uniform structure of the equal defocus zone μ_{mn}, see Algorithm 2 from Ref. 14.

3.2 Results

As a proof-of-concept, the proposed algorithm was implemented in MATLAB. The algorithm performance depends on a number of parameters, including the maximum defocus strength, the width of the equal-defocus zones and their total number, noise parameters, and the Fourier spectrum of the input object. A detailed analysis of the performance is beyond the scope of this conference paper and is a subject of a research to be published soon. In this section, we demonstrate the algorithm behaviour for a typical set of parameter for illustration purposes only.

As the ideal object, we have used an image obtained by a commercial quality inspection tool working in the debug mode of a perfectly aligned wafer (see Fig. 3, left panel). To obtain input LR images, the object image was downsampled and a random defocus blur with amplitude changing in horizontal or vertical directions was added. One of the realizations of the input images is shown in the middle panel of Fig. 3. The size of the high-resolution image was 1024×1024, down-sampling factor $\tau = 2$. An effective width of equal-defocus band was 16 pixels, with 64 bands per image in total.

The right panel of Figure 3 show the result of SANDR algorithm, 30 iterations, and the left panel of Figure 4 show the residual restoration error. The total RMS of the restoration error for this input is about 2.7%, and the total computation time was 125 seconds on i7 2.7 GHz processor.

As, to our knowledge, this is the first algorithm to address the SIND problem, we can only demonstrate its advantages over other SR methods neglecting the defocus effects. Figure 4, right panel, shows the convergence speed in terms of the residual RMS error realized for a standard SR algorithm (Projected Gradients, PG) as per Ref. 12; Sequential Minimization (SM) realized by authors using their NDR algorithm;[14] and SANDR. To speed up the calculations, the plot was generated for an HR image size of 512×512 pixels.
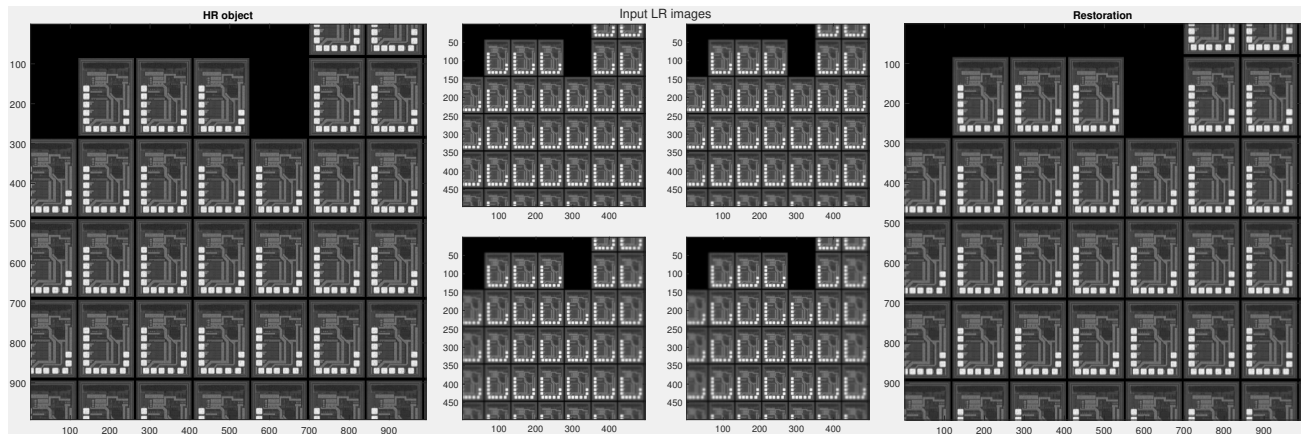
Figure 3. Original high-resolution object (left), four subpixel-shifted and blurred low-resolution images (middle) and the result of restoration of the high resolution object by SANDR algorithm after 30 iterations. One of 100 random realizations of the non-uniform blur parameter is shown.
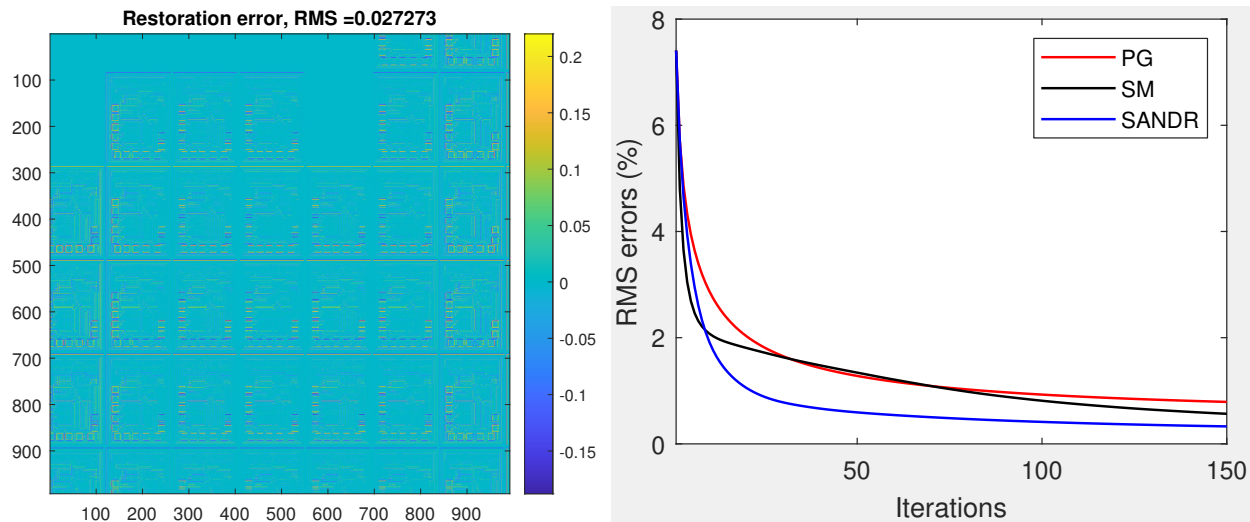

Figure 4. (Left) The residual error of the HR restoration shown in Fig. 3. (Right) RMS errors of the SR images obtained by PG, SM and SANDR as function of the iteration. The plot is calculated for an input HR object size of 512×512 pixels.

4. DISCUSSION AND CONCLUSION

The first results demonstrate stable convergence of the algorithm, and uniformity of the residual error over the image area, which, together with the low RMS of the restoration results even for a small number of iterations look promising. However there are several issues that need to be investigated further before any practical realization of the proposed method on a hardware platform. Below we outline several of them and propose possible solutions.

First of all, even with its fast convergence in terms of iterations, the algorithm is still computationally expensive. The total run time is about 2 minutes for 1024×1024 (and about 30 seconds for 512×512) image size. This is comparable with the wafer alignment time of a semiconductor inspection tool (which is about 30 seconds for a full alignment). One of the possible solutions for this issue is parallelized code and computations using GPU.

Second, the algorithm assumes that the blur parameters are known. While indeed the rough estimates for the blur size variation across the image can be quickly estimated, the prior knowledge on the problem geometry (imaging of a flat object) should be used for their automatic tuning.

Finally, the residual error in Fig. 4 is clearly localized around the edges of the chip geometry. This is a known issue of deconvolution-related problems which can be addressed by, for instance, introduction of an additional regularization term (for instance, total variance) in Eq. (13).

To conclude, we have considered an approach of increasing of the wafer inspection throughput, which proposes to increase the field of view of the camera by using smaller magnification and to computationally restore the sampling rate from several sub-pixel shifted image. To address the issue of a misalignment of the wafer image with the camera plane which might appear in result of the introduced shifts, we have developed a novel algorithm SANDR. SANDR is able to restore the high-resolution image from several low-resolution images deteriorated by a non-uniform defocus blur. We have presented the first promising results of the algorithm and outlined the main research directions which are necessary for successful implementation of the approach on a hardware demonstrator.

FUNDING

REFERENCES

[1] Huang, S., Sun, J., Yang, Y., Fang, Y., and Lin, P., "Multi-frame super-resolution reconstruction based on gradient vector flow hybrid field," *IEEE Access* **5**, 21669–21683 (2017).

[2] Zhou, Q., Chen, R., Huang, B., Liu, C., Yu, J., and Yu, X., "An automatic surface defect inspection system for automobiles using machine vision methods," *Sensors* **19**(3) (2019).

[3] Camtek Ltd., "Camtek Inspection and metrology, https://www.camtek.com/solution/front-end/," (2022).

[4] Yaroslavsky, L. P., [*Advances in Sampling Theory and Techniques*], SPIE ("2020").

[5] Irani, M. and Peleg, S., "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing* **53**(3), 231–239 (1991).

[6] Ur, H. and Gross, D., "Improved resolution from subpixel shifted pictures," *CVGIP: Graphical Models and Image Processing* **54**(2), 181–186 (1992).

[7] Kim, S., Bose, N., and Valenzuela, H., "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **38**(6), 1013–1027 (1990).

[8] Kim, S. and Su, W.-Y., "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans Image Process.* **2**(4), 534–542 (1993).

[9] Sauer, K. and Allebach, J., "Iterative reconstruction of bandlimited images from nonuniformly spaced samples," *IEEE Transactions on Circuits and Systems* **34**(12), 1497–1506 (1987).

[10] Sroubek, F., Cristobal, G., and Flusser, J., "A unified approach to superresolution and multichannel blind deconvolution," *IEEE Transactions on Image Processing* **16**(9), 2322–2332 (2007).

[11] Li, X., Hu, Y., Gao, X., Tao, D., and Ning, B., "A multi-frame image super-resolution method," *Signal Processing* **90**(2), 405–414 (2010).

[12] Farsiu, S., Robinson, D., Elad, M., and Milanfar, P., "Fast and robust super-resolution," in [*Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*], **2**, II–291 (2003).

[13] Yamauchi, Y., Yamashita, J., Fukui, Y., Yokoyama, K., Sekiya, T., Ito, E., Kanai, M., Fukuyo, T., Hashimoto, D., Iseki, H., and Takakura, K., "A dual-view endoscope with image shift," in [*CARS 2002 Computer Assisted Radiology and Surgery*], Lemke, H. U., Inamura, K., Doi, K., Vannier, M. W., Farman, A. G., and Reiber, J. H. C., eds., 183–187, Springer Berlin Heidelberg, Berlin, Heidelberg (2002).

[14] Thao, N. H., Soloviev, O., Noom, J., and Verhaegen, M., "Nonuniform defocus removal for image classification," Manuscript submitted to IEEE Transactions on Image Processing in February 2021.

[15] Verhaegen, M. and Verdult, V., [*Filtering and System Identification: A Least Squares Approach*], Cambridge University Press (2007).

[16] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009).

[17] Monga, V., [*Handbook of convex optimization methods in imaging science*] (2017).