A Super-Resolution Model for Improving the Precision of Wafer Mark Alignment

Sheng Lei1,a*, Sen Lu2,b, Kaiming Yang3,c

1 State Key Laboratory of Tribology Department of Mechanical Engineering, Tsinghua University, Beijing, China
2 State Key Laboratory of Tribology Department of Mechanical Engineering, Tsinghua University, Beijing, China
3 State Key Laboratory of Tribology Department of Mechanical Engineering, Tsinghua University, Beijing, China

aemail: leis18@mails.tsinghua.edu.cn, bemail: lusen@mail.tsinghua.edu.cn, cemail: yangkm@tsinghua.edu.cn

Abstract. Wafer bonding machine uses industrial camera to recognize marks on top wafer and bottom wafer, compute deviation, and move top wafer to align the bottom wafer. The alignment precision mainly depends on the camera resolution, high resolution industrial camera is expensive, while classical image up-sampling methods such as bicubic interpolation don’t have good effect. To improve the alignment precision, a super-resolution model is proposed. Main component of this model is convolutional neural network. The first two convolutional layers are to extract feature on wafer image, the next convolutional layer is used for nonlinear mapping, and the final one outputs super-resolution image. Peak Signal to Noise Ratio (PSNR) is used to evaluate the similarity of super resolution image and the target high resolution image. It’s proved by experiments that the super-resolution model has better effect than classical image interpolation methods. This research result can also be applied to other equipment using industrial cameras.

1. Introduction
Recently, as image sensor develops rapidly, control technology that uses image processing systems and techniques has been widely used in various industrial fields. For example, Wermes [1] used pixel detection to solve micro area problems, and Miyasaka [2] applied pixel-based method to make area classification. Besides, wafer bonding is also one field that mainly uses image processing techniques.

In the field of wafer bonding, accurate alignment is necessary to ensure wafer bonding quality, which directly influences the fabrication of advanced semiconductor components. To achieve wafer-to-wafer alignment, lots of alignment techniques have been proposed during the last decades. Among them an alignment technique named smart view alignment uses two pairs of industrial cameras to recognize alignment marks of top and bottom wafers, then calculates deviation of top marks and bottom marks and compensates it. Fig. 1 and Fig. 2 illustrate how the smart view alignment works. Another mainstream alignment technique is based on the observation of transmission images of the concentricity...
of top and bottom alignment marks [3]. Both of these alignment methods are fundamentally limited by the resolution of wafer mark image. Therefore, more precise alignment techniques need to be developed.

In recent years, there were many researches on how to improve wafer alignment accuracy. Park [4] proposed a system that spun the wafer one 360 degree and measured the edge position to calculate deviation. Kim [5, 6] proposed a method for constructing a general alignment model and formulating equations from three coordinates based on the assumption of vision detection and precise position control in 2006, then developed an iterative automatic alignment algorithm in 2007, which was also based on accurate vision detection. Liudi Jiang [7] proposed a passive approach which exploited mechanical principles of kinematic and elastic averaging to achieve nano-precision bonding alignment. Chenxi Wang [8] proposed a novel wafer-to-wafer alignment method using centrosymmetric moiré gratings. Ming-Fei Chen [9] developed pattern matching identification and an algorithm to obtain the crisscross mark positions of two wafers’ coordinates for the wafer bonding process. Joeri De Vos [10] discovered that wafer-to-wafer alignment and bonding process introduced nonlinear distortion, which had an impact on last alignment. By taking into account the described alignment tolerances, the alignment accuracy can be improved. An unusual way to improve alignment accuracy is to increase the resolution of wafer images through algorithms, so as to allow image processing algorithms calculating deviation from the top marks and bottom marks more accurately. In computer vision domain, Ji Linlin [11] proposed a bicubic interpolation image magnification method based on local interpolation, every new pixel interpolated in the image is calculated by neighborhood weighted average. Chao Dong [12]
proposed an image super-resolution method using deep convolutional networks, which can output high quality reconstructed images. In recent research, residual block is proposed to decrease complexity of network and prevent vanishing gradient problem when the network is deep, it is applied to image classification and object detection domain widely.

In this paper, we introduce a super-resolution model that increases resolution of the wafer mark image. The model is combined of three parts, the first part extracts wafer mark image from the original image, the second part is based on bicubic interpolation that regularizes different image size, the third part is a four-layer convolutional neural network, which reconstructs low-resolution image to high-resolution image. Back propagation algorithm is used to train the neural network, and PSNR is applied to evaluate the super-resolution effect. Chapter 2 introduces the structure of super-resolution model, chapter 3 gives experiment condition, results and discussion. In chapter 4, we give conclusion of this paper. Chapter 5 is acknowledgement.

2. Structure of super-resolution model
In this paper, the proposed super-resolution model is combined of three parts. The first step is to detect and extract wafer mark. The second step is image magnification, which is based on bicubic interpolation. The third step is pixel reconstruction, which inputs the coarse magnified image into a convolutional neural network to improve pixel quality.

2.1. Wafer mark detection and extraction
The original wafer image has 3840*2748 pixels, which is too large for vision algorithms to process. Because wafer mark is a very small part of the picture, handling it alone can improve speed and accuracy. So we first detect and extract the wafer mark part by pattern matching.

2.2. Image magnification
Classical image magnification algorithms include nearest-neighbor interpolation, bilinear interpolation, bicubic interpolation, etc. Nearest-neighbor interpolation directly uses the nearest pixel value to be the value of the inserted pixel, which is the fastest algorithm but has the worst effect. Bilinear interpolation calculates the inserted pixel value based on the weighted average of the pixel values of 4 points closest to the inserted point, which is slower than the nearest-neighbor interpolation, but has better effect. Bicubic interpolation calculates the inserted pixel value based on the weighted average of the pixel values of 16 points closest to the insertion point, which always produces more effective and accurate interpolation images. Fig. 3 shows the principle of bicubic interpolation.

In this paper, we use an improved bicubic interpolation algorithm to magnify the original image. The principle of bicubic interpolation algorithm can be seen in Figure 3. The pixel value of P can be solved by the following expression:

\[ P = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} W(i)W(j) \]  

(1)

where \( P \) represents the inserted pixel value, \( a_{ij} \) represents the original pixel in row \( i \) and column \( j \), \( W(i) \) represents the row weight and \( W(j) \) represents the column weight. So image magnification can be represented by the following expression:

\[ X = B(L) \]  

(2)

where \( L \) represents the original low-resolution image, \( B() \) represents the bicubic interpolation function, and \( X \) represents the magnified image. Because new pixels in \( X \) is calculated by weighted average, quality of new pixel is not high, and the edge of the magnified image is blurred.
2.3. Pixel reconstruction

We use convolutional neural network to reconstruct interpolated pixels in the magnified image, structure of this network is illustrated in Fig. 4, which consists of four convolutional layers and denotes four steps:

- **Low-level feature extraction:** this operation extracts fundamental geometric features from the magnified image $X$, and uses feature map to represent these features, makes it easier to extract high-level feature.

- **High-level feature representation:** this operation extracts high-level features from feature maps extracted by convolutional layer 1, makes it easier to map low-quality pixels to high-quality pixels. These high-level features are represented as high-dimensional vectors.

- **Nonlinear mapping:** this operation maps the high-dimensional vectors which represent high-level features to another vector space nonlinearly, thus can learn some nonlinear relationships during the training process.

- **Interpolated pixel reconstruction:** this operation directly generates every reconstructed pixel, and the output image is expected to be similar to the ground truth image.

$$C_1(X) = \max (B_1 + X \ast W_1, 0)$$

Fig. 4. Structure of the pixel reconstruction neural network. It uses convolutional layer as the output, which is different from most neural networks.

2.3.1. Low-level feature extraction. Traditional neural network uses matrix multiplication to realize the mapping from input to output, each parameter in the coefficient matrix interacts with each input unit separately, which causes the coefficient matrix being large, complicated, time-consuming and difficult training. As for convolutional network, the convolutional kernel is far smaller than the input size, which is characterized by sparse interaction, can greatly improve the efficiency of the algorithm and play an important role in detecting small fundamental features in the whole image. In our formulation, the first layer is expressed as the following operation $C_1(X)$:

$$C_1(X) = \max (B_1 + X \ast W_1, 0)$$

where $W_1$ represents the weight matrix, $B_1$ represents the bias matrix, activation function is Relu, and ‘$\ast$’ denotes the convolution operation. In this case, we use a 5*5*64 kernel.

2.3.2. High-level feature representation. This step is the same as the first step, but it extracts high-level features, so the kernel size is larger than the first layer. For most situations, increasing network
depth gains better effect than increasing kernel size. In our formulation, the second layer is expressed as the following operation $C_2(X)$:

$$C_2(X) = \max (B_2 + C_1(X) \ast W_2, 0)$$

(4)

where $W_2$ denotes the weight matrix, $B_2$ denotes the bias matrix, activation function is Relu. In this case, we use a $7 \times 7$ kernel, which is larger than the first kernel. The reason to choose $7 \times 7$ kernel is that it can scan more abstract features.

2.3.3. Nonlinear mapping. This step maps the vectors generated by (4) to another vector space, and only $1 \times 1$ kernel has valid interpolation for nonlinear mapping. The third layer can be expressed as the following operation $C_3(X)$:

$$C_3(X) = \max (B_3 + C_2(X) \ast W_3, 0)$$

(5)

where $W_3$ denotes the third layer’s weight matrix, $B_3$ denotes the third layer’s bias matrix.

2.3.4. Interpolated pixel reconstruction. The last operation is interpolated pixel reconstruction. In general, the last layer in neural network is connection layer. For most situations such as object detection and image classification, connection layer is useful. But in this situation, the output should be image pixel values, so we define a convolutional layer as the final layer, which can output super-resolution wafer mark image. This reconstruction layer can be expressed as the following operation $C_4(X)$:

$$C_4(X) = \max (B_4 + C_3(X) \ast W_4, 0)$$

(6)

where $W_4$ represents the last layer’s weight matrix, $B_4$ represents the last layer’s bias matrix. Kernel size of this filter is $5 \times 5 \times 1$, for both input image and output image are grayscale.

2.4. Model training

We use mean square error as loss function of this model, while back propagation algorithm is applied to train the convolutional neural network. By calculating the mean square error between the output of the super-resolution model and the corresponding high-resolution image label, and taking it as the objective function, weight matrix and bias matrix that minimize the objective function can be solved. Specifically, gradient is calculated by the small-batch stochastic gradient descent algorithm, the weight matrix and bias matrix are updated according to the learning rate, and the local optimal values are obtained by multiple iterations. The mean square error loss function can be expressed as:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \| C_4(X; \theta) - Y_{label} \|^2$$

(7)

where $\theta$ denotes all the parameters such as weights and bias of different layers.

3. Results & Discussion

In this paper, the super-resolution model is mainly designed for wafer mark image preprocessing, so the experiments are designed based on wafer mark alignment. First, we use high-resolution industrial camera to take some original high-resolution images, and these high-resolution images are extracted by a $320 \times 256$ pixels’ filter to find the rough position of wafer mark. Then these high-resolution $320 \times 256$ images are down-sampled by many different kinds of methods such as nearest neighbor interpolation, discarding pixels evenly and so on, these down-sampled images are called low-resolution image and have $160 \times 128$ pixels. After that, these low-resolution images are magnified by improved bicubic interpolation algorithm, then they are put into the pixel reconstruction neural network, and finally the model outputs super-resolution image which also has $320 \times 256$ pixels. By calculating PSNR of the super-resolution image and original high-resolution image, we can evaluate effect of the super-resolution model. PSNR is calculated by the following formulas:

$$PSNR = 20log_{10}(\frac{256}{\text{MSE}})$$

(8)
\[ MSE = \frac{1}{320 \times 256} \sum_{i=0}^{255} \sum_{j=0}^{255} ||I(i,j) - K(i,j)||^2 \]  

(9)

from the results shown in Table 1, it’s obvious that the proposed super-resolution model can generate high resolution images that have better quality than classical interpolation methods.

Table 1. Comparing effects of super-resolution model and bicubic interpolation by PSNR. 45 experiments have been taken and here are some results of them.

| Test Number | PSNR of bicubic interpolation(dB) | PSNR of super-resolution model(dB) |
|-------------|-----------------------------------|------------------------------------|
| 1           | 83.615                            | 88.639                             |
| 2           | 83.508                            | 88.659                             |
| 3           | 84.014                            | 90.757                             |
| 4           | 84.020                            | 89.943                             |
| 5           | 85.097                            | 91.626                             |
| 6           | 85.851                            | 92.135                             |
| 7           | 82.176                            | 88.582                             |

Fig. 5. (a) is a low-resolution wafer image, (b) is generated by the proposed model, (c) is generated by bicubic interpolation, and (d) is high-resolution image which is used as a comparison.

Fig. 6. (a) is extracted from image generated by bicubic interpolation, (b) is extracted from image generated by the proposed model. It’s obvious that (b) has better quality than (a).

Besides the comparison of PSNR, Fig. 5 shows the high-resolution image, super-resolution image, image processed by classical bicubic interpolation, and low-resolution image, while Fig. 6 is a partially enlarged view. It’s clear that super-resolution image has better quality than low-resolution...
image and classical interpolation image. The final alignment accuracy mainly depends on the image accuracy, while image accuracy depends on the size that one pixel exactly is. Distance between two adjacent pixels in the low-resolution image is 3.334μm. Using this super-resolution model, the distance can be decreased to 1.667μm. In other word, this super-resolution model can make a low-resolution industrial camera works as well as a high-resolution one. As the super-resolution model has neural network structure and good generalization ability, it can also be used to improve the resolution of high-resolution industrial camera, thus achieving sub-pixel accuracy.

4. Conclusion
In this paper, we proposed a super-resolution model that aimed to improve the resolution of wafer mark image, thus made it easier to detect mark and calculate its accurate position, and finally improved the precision of mark alignment. Experiments showed that with a magnification factor of 2*2, the super-resolution model had better effect than classical bicubic interpolation algorithm. Further work needs to be done to improve the magnification factor and super-resolution quality, and finally improve the mark alignment accuracy. This technology can also be applied to other industrial machines that use industrial cameras.

Acknowledgement
This research was supported by State Key Laboratory of Tribology Tsinghua University grants SKLT2020D23. We thank for the fund.

References
[1] Wermes, N. (2013) Pixel detector for particle physics and imaging application. Nucl. Instrum. Methods Phys. Res., A, 52: 277–288.
[2] Miyasaka, T., Okuro, T., Zhao, X., Takeuchi, K. (2016) Classification of Land Use on Sand-Dune Topography by Object-Based Analysis, Digital Photogrammetry, and GIS Analysis in the Horqin Sandy Land, China. Environments, 3(17).
[3] Jongwon K. (2018) New Wafer Alignment Process Using Multiple Vision Method for Industrial Manufacturing. Electronics., 7(39).
[4] Park, H.L., Lyou, J. (2002) A Wafer Alignment Method and Accuracy Evaluation. J.Inst. Control Robot. Syst., 8: 812-817.
[5] Kim, H.T., Song, C.S., Yang, H.J. (2006) Algorithm for automatic alignment in 2D space by object transformation. Microelectron. Reliab., 46: 100-108.
[6] Kim, H.T., Yang, H.J., Baek, S.Y. (2007) Iterative algorithm for automatic alignment by object transform. Microelectron. Reliab., 47: 972-985.
[7] Jiang, L., Pandraud, G., French, P.I., Spearing, S.M., Kraft, M. (2007) Wafer Bonding with Nanoprecision Alignment for Micro/Nano Systems. In: International Solid-State Sensors, Actuators and Microsystems Conference. Lyon, France. pp. 2103-2106.
[8] Wang, C., Taniyama, S., Wang, Y., Suga, T. (2008) High-Precision Alignment for Low Temperature Wafer Bonding,” ECS Transactions, 16(8): 539-548.
[9] Chen, M., Ho, Y., Wang, S. (2010) A Fast Positioning Method with Pattern Tracking for Automatic Wafer Alignment. In: International Congress on Image and Signal Processing. Yantai. pp. 1594-1598.
[10] De Vos, J., Peng, L., Phommahaxay, A., Van Ongeval, J., Miller, A., Beyne, E., Kurz, F., Wagenleiter, T., Wimpling, M., Uhrmann, T. (2016) Importance of alignment control during permanent bonding and its impact on via-last alignment for high density 3D interconnects. In: IEEE International 3D Systems Integration Conference. San Francisco. pp. 1-5.
[11] Ji, L., Wang, P., Zhang, Y. (2019) Bicubic image magnification based on local interpolation. Journal of Graphics, 40(1): 143-149.
[12] Dong, C., Chen, C.L., He, K., Tang, X. (2016) Image Super-Resolution Using Deep
Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2): 295-307.