Image-Based Parametric Pattern Recognition for Micro- and Nano- Defect Detection

Sergey Belikov*. Chanmin Su**. Marian Enachescu***.

* SPM Labs LLC, Tempe, AZ, USA; ** Shenzhen Acad. of Robotics, China; *** Politehnica Univ. of Bucharest, Romania

Abstract: Detection and characterization of defects at micro- and nano-scale is one of the most important tasks of the production process control technologies at these scales. Examples include flat panel displays (FPD) and MEMS (defects at micro scale), and hard disk drive substrates up to 95 mm in diameter (defects at nano scale). Although different measurement technologies are applied, such as infrared thermography at micro-scale and Atomic Force Microscopy (AFM) at nano-scale, a similar technique of image-based defect recognition and location can be utilized, namely parametric pattern recognition, where the defect is defined by a specific set of inequalities in the space of selected measurable variables. We describe the technique and demonstrate it for the following applications: 1) mechanical defects (high friction) in MEMS; 2) electrical defect locations (shorts, opens) on FPD; and 3) nano-asperities on hard disk drive surfaces. The first two use infrared thermography, and the third uses multi-channel AFM scanning. Both large and small fields of view images are needed for analysis.

Keywords: Micro- and Nano-Mechatronics Systems, Defect Detection, Pattern Recognition, Modeling

1. INTRODUCTION

In the last few decades, micro- and nano-scale technologies made a significant progress in further increasing density of electrical and mechanical components. This progress, however, requires much stricter demands on defect identification, including overheating due to mechanical friction, electrical shorts and opens, debris of nanometer sizes, etc. At these small sizes, no direct measurements is possible, and techniques such as Infrared Thermography at micro-level for estimation of temperature or current, and Atomic Force Microscopy (AFM) at nano-level for locating nano-debris must be utilized.

In this paper we describe some computational details behind patented test technologies in manufacturing of MEMS [Muhlstein, C.L., Brown, S.B., and Ritchie, R.O. (2001)] and [Sharpe, N. Jr. (2002)]; Flat Panel Displays [Enachescu, M., and Belikov, S. (2005)] and [Belikov, S., and Enachescu, M. (2009)]; and disks for data storage [Belikov, S., Shi, J., and Su, C. (2008)] and [Su, C., and Belikov, S. (2011)].

2. INFRARED THERMOGRAPHY FOR RESONANT STRUCTURE TESTS

The resonant structure used in the test of this section was described in [Sharpe, N. Jr. (2002): Fig. 3-10]. One of its applications can be found in [Muhlstein, C.L., Brown, S.B., and Ritchie, R.O. (2001): Fig. 1]. In this section we use the Infrared camera (FLIR Systems, Inc.) with high resolution shown in Fig. 1, on the right. We use some characteristics of this camera to measure temperatures based on infrared images. The infrared images of the structure under test [Muhlstein, C.L., Brown, S.B., and Ritchie, R.O. (2001)] are shown in Fig. 2. In section 2.1 we estimate the temperature

experimentally by analysing the infrared images in Fig. 2, and in section 2.2 we calculate the temperature from theoretical model based on mechanics and heat transfer shown in Fig. 3. We will compare the results to justify both methods and use the infrared thermography for defect detection in section 4.

Fig. 1. Two Mid-Wavelength InfraRed cameras (Merlin Mid InSb MWIR) with the large field of view (left) and the small one (right) with resolution 7.57[μm/pixel]. Cameras can be moved vertically for changing fields of view and autofocusing.

2.1 Temperature Difference Estimation based on Infrared Images

Infrared images with actuated and non-actuated comb drive built in the MEMS system, as shown in Fig. 2.a and 2.b, are very close and the discrepancy can be only seen in their difference. However, the difference of raw images also does not provide reliable information about the temperature difference because it is obscured by noise comparable with the difference. Images should be processed before calculating the difference. Averaging sequences of images and filtering reduces the noise and generates reliable information about temperature differences. Effects of nonlinear filtering on image enhancement are described in [Gonzalo, L., Paredes, L., and Mullan, J. (2000)]. Median filter of size 3x3 has been used for this application. This filter effectively eliminated bad pixels and smoothed images. The details of the calculation that enhanced the differential image, as shown in Fig. 2.c, can be found in [Enachescu, M., and Belikov, S. (2005)].

A conventional model of the output of the infrared camera is described in [Sparrow, E.M., and Cess, R.D. (1978)] and [Belikov, S., et al. (1995)]. Simplified for constant total emissivity ε, total responsivity R_{total}, and assuming Stefan-Boltzmann law, a theoretical approximation of the output of the infrared camera is

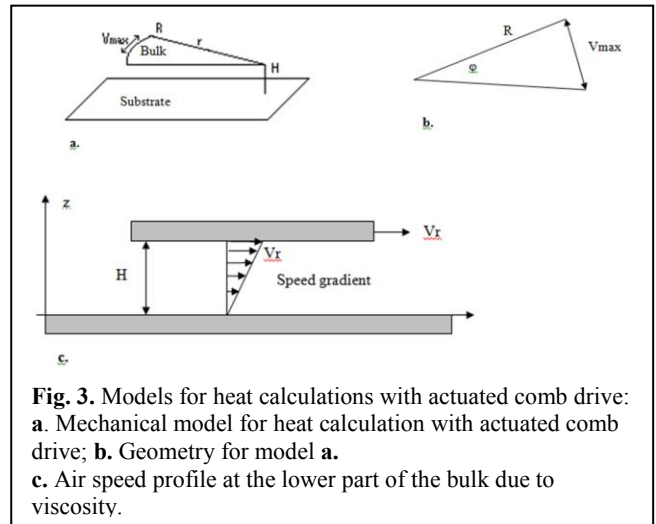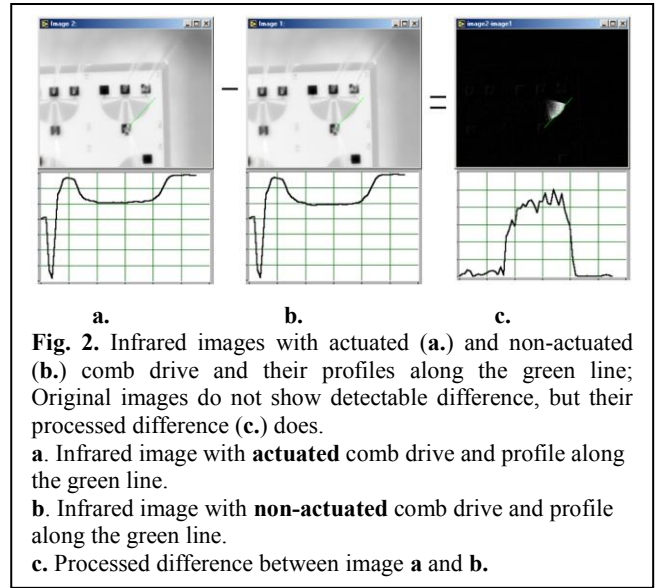$$S(T) = \varepsilon R_{\text{total}} F(T), \quad F(T) = \sigma T^4 \qquad (1)$$

where

- $S(T)$ [units] is the output of the camera in [units], i.e. in pixel values of the whole (12 bit) range ($[-2^{11}, +2^{11}-1]$ for this camera);

- ε is total emissivity of silicon (assume $\varepsilon \approx 0.06$) [Timans, P.J. (1995)];

- T [K] is absolute temperature;

- R_{total} [units/(W/m^2)] is the change of infrared image value as a result of 1[W/m^2] change of radiation energy from unit area of the target black-body source;

- $F(T)$ [W/m^2] is a radiation law;

- $\sigma \approx 5.67 \cdot 10^{-8}$[W/(m^{2}K^4)] is the Stefan-Boltzmann constant.

Pixel values in Fig. 2.a and 2.b are equal to $S(T+\Delta T)$ and $S(T)$ respectively, $T \approx 20^\circ$C, and ΔT is the difference between temperatures of the surface with and without comb drive actuation. For small ΔT

$$\Delta S = \varepsilon R_T \Delta T, \quad R_T\left[\text{units/K}\right] = R_{\text{total}} \cdot dF(T)/dT$$

where

- R_T [units/K] is the change of infrared image value as a result of 1°C temperature change of the target black-body source, called responsivity of the infrared camera. For the camera used in this experiment $R_T \approx 1275$ [units/K].

Fig. 2. Infrared images with actuated (**a.**) and non-actuated (**b.**) comb drive and their profiles along the green line; Original images do not show detectable difference, but their processed difference (**c.**) does.
a. Infrared image with **actuated** comb drive and profile along the green line.
b. Infrared image with **non-actuated** comb drive and profile along the green line.
c. Processed difference between image **a** and **b**.

Fig. 3. Models for heat calculations with actuated comb drive:
a. Mechanical model for heat calculation with actuated comb drive; **b.** Geometry for model **a.**
c. Air speed profile at the lower part of the bulk due to viscosity.

The procedure to measure ΔT was the following. The neighbour points were selected at the images in Fig. 2 and pixel values were measured for both (processed) images with actuated and non-actuated comb drive respectively. The results are summarized in the table below.

Table 1. Infrared measurements at 4 points

Point #	Pixel values at the image with comb drive		Difference ΔI
	non-actuated	actuated	
1	19,455	20,098	643
2	19,134	20,088	954
3	19,230	20,116	886
4	19,259	20,110	851
Av.	19,269	20,103	834

The average difference $\Delta I=834$ is the difference of image values scaled by the camera range $[-2^{11}, 2^{11}-1]$. Infrared image differences in [units] relates to ΔI by the formula

$$\Delta S = \Delta I \cdot (I_{\max} - I_{\min})/(2^{12}-1)$$

where $[I_{\max}, I_{\min}]$ is the selected range for current image acquisition. For this application
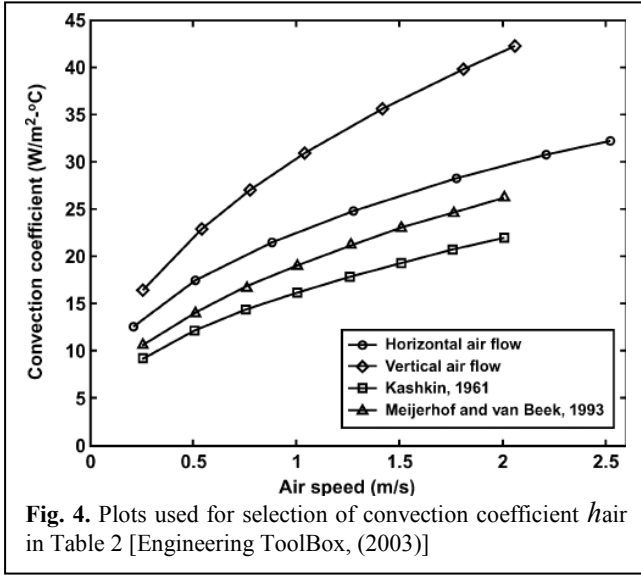
Fig. 4. Plots used for selection of convection coefficient $h$air in Table 2 [Engineering ToolBox, (2003)]

$$I_{\max} - I_{\min} = 1000 \quad \Rightarrow \quad \Delta S \approx 204[\text{units}]$$

and, with $\varepsilon \approx 0.06$, estimated temperature difference is

$$\Delta T = \Delta S / \varepsilon R_T \approx 2.7[K] \qquad (2)$$

In this derivation we ignored the image intensity loss due to the optical lensing system between the image and Focal Plane Array.

2.2 Temperature Difference Calculation based on Heat Balance Model

This model assumes the configuration shown in Fig. 3. Bulk at distance H from the substrate vibrates in the way that the points of radius R have maximum linear speed $V_{\max}$. The parameters used in the model are presented in Table 2.

Table 2. Parameters of the heat balance model

Parameter	Notation	Value	Comments
Viscosity of air	μ_{air}	$1.8 \cdot 10^{-5}\left[\frac{kg}{m \cdot s}\right]$	At 20°C [Thomas, L.C. (1993): p. 833]
Convection coefficient of air	h_{air}	$30-40\left[\frac{W}{m^2 K}\right]$	For small forced convection: see Fig.4 [Engineering ToolBox, (2003)]
Distance between bulk and substrate	H	$2[\mu m]$	See Figure 3
Length of bulk	R	$300[\mu m]$	See Figure 3
Max speed of vibrating edge of the bulk	$V_{\max}$	$5[m/s]$	See Figure 3
Total emissivity of silicon at 20°C	ε	0.06	For 20°C: [Timans, P.J. (1995): p. 82, Fig. 30]

Let φ [rad] be the maximum angle of vibrating bulk; R [m] the length ("radius") of the bulk as shown in Fig. 3.a, 3.b; and v [Hz] the frequency of vibration. The following relationship is true

$$\varphi = V_{\max} / \pi R v \qquad (3)$$

To prove this relationship, we calculate the average speed $\overline{V}_R$ of the end of the bulk during the period of vibration:

$$\overline{V}_R = V_{\max} \frac{1}{\pi} \int_0^\pi \sin t\, dt = \frac{2V_{\max}}{\pi} \qquad (4)$$

Another formula for the average speed is

$$\overline{V}_R = 2\varphi R / T = 2\varphi R v \qquad (5)$$

where T is the period of the vibration. Relationship (3) follows from these two formulas.

Figure 3 c. shows a conventional (Newton) model of air speed distribution due to viscosity. Shown is the cross-section of beam-substrate system at the distance r from the center of rotation. Friction force $F_{fr}(r)$ acting on a small area S at this location is

$$F_{ff}(r) = \mu_{air} |\text{grad } v_r| S \qquad (6)$$

where μ_{air} [kg/(m·s)] is the viscosity of air and

$$\text{grad } v_r = \frac{dv_r}{dz} = \frac{V_r}{H} = \frac{V_R r}{HR} \qquad (7)$$

Substituting this expression to (6) yields

$$F_{ff}(r) = \mu_{air} \frac{V_R r}{HR} S \qquad (8)$$

Work made against this friction force during the cycle of vibration is

$$A_{ff}(r) = \overline{F}_{ff}(r) 2\varphi r = 2\varphi \mu_{air} \frac{\overline{V}_R r^2}{HR} S \qquad (9)$$

where over-score denotes average value over the period of vibration. The power dissipation during the period is

$$W(r) = \frac{A_{ff}(r)}{T} = A_{ff}(r) v = 2\varphi \mu_{air} \frac{2V_{\max}}{\pi} v \frac{r^2}{HR} S \qquad (10)$$

Substituting φ from (3) yields

$$W(r) = \frac{4\mu_{air}}{\pi^2 H} V_{\max}^2 \left(\frac{r}{R}\right)^2 S \qquad (11)$$

Formula (11) is the power dissipation from the lower side of the beam due to air viscosity. When calculating heat balance, only this power dissipation due to viscosity and the power due to convection from the upper side of the beam are taken into account. The following factors are neglected: air viscosity at the upper side of the beam (because the gradient is significantly higher at the lower side where air is "pulled" by two surfaces at 2μm from each other); convection from the lower part of the beam (because this convection is prevented by the small distances between the surfaces and viscosity of the air); heat conduction in xy direction (because period T is small); immediate heat transfer is assumed in z-direction (because of small thickness of the beam); influence of comb drives (electro-static actuator and capacitive sensor) – this is actually causes some non-symmetry of infrared images, but may be neglected at central radial line.

Thus, we assume the energy conservation in the form: [Power Dissipation from Friction] = [Power Loss by Convection], i.e.

$$\frac{W(r)}{S} = h_{air}\left(T(r) - T_{air}\right) \qquad (12)$$

where $T_{air} \approx 20^\circ C$ is the temperature of surrounding air. Substituting $W(r)$ from (11) yields

$$\frac{4}{\pi^2}\mu_{air}\frac{V_{max}^2}{H}\left(\frac{r}{R}\right)^2 = h_{air}\left(T(r) - T_{air}\right) \qquad (13)$$

Then the formula for temperature difference at radial position R is

$$\left(T(R) - T_{air}\right) = \frac{4}{\pi^2}\frac{\mu_{air}}{h_{air}}\frac{V_{max}^2}{H} \qquad (14)$$

Calculations by this formula with the parameters from Table 2 are presented in Table 3.

Table 3. Calculations by Formula(14)

h_{air} [W/m^{2}K]	$T(R)$-20$^\circ C$ [K]
30	3.0
35	2.6
40	2.3

The table is in agreement with formula (2). This agreement may be considered as a reasonable justification of both methods: experimental based on infrared imaging and theoretical based on simplified heat balance.

2.3 Infrared Test of Acceptable Friction for Resonant Structures

Theoretically normal temperature rise due to the friction was calculated and confirmed by infrared measurements. This allows us to design a test procedure for acceptable friction: 1) measure the temperature rise at the actuated beam location; 2) compare the measurement with nominal interval, e.g. [2.3°C, 3.0°C] suggested by Table 3. The test is passed if the measured temperature is within the nominal interval.

This test is relatively easy because the location of the beam is known a priori. More challenging is the defect location (e.g. electrical shorts at FPD pixels) that requires pattern recognition and will be described in section 4 after we review the parametric technique in section 3.

3. IMAGE BASED PARAMETRIC PATTERN RECOGNITION

This section presents detailed descriptions of the framework briefly outlined in [Belikov, S., Shi, J., and Su, C. (2008)]. The framework will be applied in section 4 to electrical shorts/opens detection using infrared thermography (described in section 2), and in section 5 to nano-asperity detection in data storage industry using multi-channel AFM images. In both applications, the defects can be easily confused with noise or image artefacts. Comprehensive understanding and definition of what a defect is allows separating defects from other singularities on the multi-channel images.
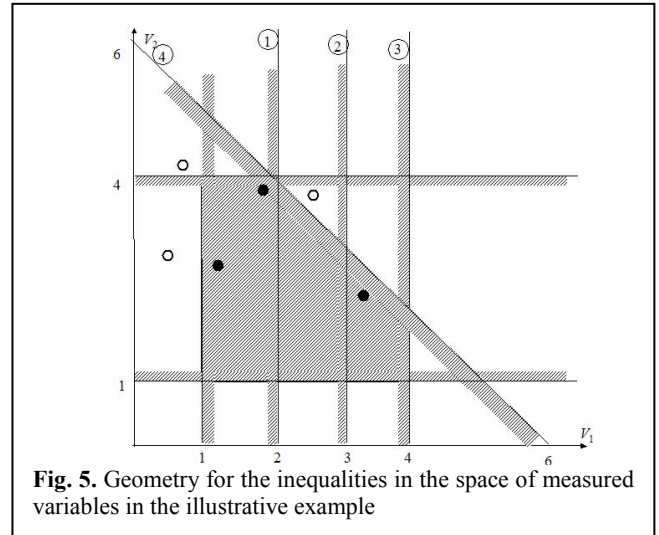
Fig. 5. Geometry for the inequalities in the space of measured variables in the illustrative example

There are two approaches to pattern recognition and location on multi-channel images: template-based and parametric-based. Template-based algorithms locate regions on the image that match an a priori reference pattern (template). Parametric-based algorithms can be applied when interested patterns are changing from image to image and cannot be characterized by an a priori image template. These changing patterns, however, can be identified by a set of measurable variables such as geometrical and regression properties that are restricted to a certain parametrized region in the measured variables' space [Duda, R., Hart, P., and Stork, D. (2001)].

The patterns in the multi-channel images are identified by a set of *thresholds* τ_i, $(i=1,..,n_\tau)$. These thresholds divide the pixels on every related image channel into two regions (with pixel values below and above the thresholds) and define the *particles* on the images, i.e., connected regions of pixels with values above the thresholds. The coordinates of the pixel with predefined location (e.g., location of the peak value inside the particle or a well-defined corner of the particle) are called a *location of the particle*. Particles appeared in different imaging channels are called *related particles* if they satisfy certain conditions (including closeness of their locations).

Besides locations, many other variables can be measured on the related particles including geometrical and regression variables. A set of related particles is called a *pattern* if a predefined system of inequalities

$$F_i(\mathbf{V}, \mathbf{P}) \le 0, i = 1, \ldots, n_F \qquad (15)$$

holds for the measured vector $\mathbf{V}$, and parameter vector $\mathbf{P}$ where $\mathbf{P}=(P_1,\ldots,Pn_P)$, n_P is the number of parameters. $\mathbf{V}=(V_1,\ldots,Vn_V)$ is the *vector of measured variables*, n_V is the number of variables. In literature functions F_is are called discriminant functions [Duda, R., Hart, P., and Stork, D. (2001)].

In an illustrative example shown in Fig. 5, there is a two-dimensional measurement vector $\mathbf{V}=(V_1, V_2)$. The shaded area can be described with inequalities $1 \le V_1 \le 4$, $1 \le V_2 \le 4$, $V_1+V_2 \le 6$. With $\mathbf{P}=(1,4,1,4,6)$ the inequality functions in (15) for this example are the following: $F_1=-V_1+P_1$, $F_2=V_1-P_2$, , $F_3=-V_2+P_3$, $F_4=V_2-P_4$, $F_5=V_1+ V_2-P_5$.

For particles in AFM images, V_1, V_2 can represent area and orientation of the particles, and parameters restrict the values of these variables. Other variables may be regression coefficients between related AFM images of different channels.

The values of parameters can be optimized using training samples. Trained parameters along with the set of inequalities (15) describe the pattern. If a set of inequalities (15) does not hold, the set of related particles is classified as a *false pattern*. A criterion of optimization (loss function) is defined as loss due to pattern recognition errors. There are two kinds of errors:

1. Classify true pattern as false (missed pattern), with α represents the loss due to missing pattern; and

2. Classify false pattern as true, with β as the loss due to false hit.

Let us assume that there are *N training samples* (i.e. set of related particles) with measured vectors $\mathbf{V}^j$, $j=1,\ldots,N$,

$$\chi\left(\mathbf{V}^j\right)=\begin{cases}1, \text{if } j^{\text{th}} \text{ sample is a true pattern}\\0, \text{if } j^{\text{th}} \text{ sample is a false pattern}\end{cases}$$

and

$$\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^j\right)=\begin{cases}1, \text{if } F_i\left(\mathbf{V}^j,\mathbf{P}\right)\leq 0, i=1,\ldots,n_F\\0, \text{otherwise}\end{cases}$$

Then the criterion of minimization is the following

$$\Phi(\mathbf{P})=\alpha\sum_{j=1}^N\chi\left(\mathbf{V}^j\right)\left[1-\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^j\right)\right]+\beta\sum_{j=1}^N\left[1-\chi\left(\mathbf{V}^j\right)\right]\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^j\right)\to\min_{\mathbf{P}} \quad (16)$$

where the first term is the loss function due to missed patterns and the second is the loss function due to false hits.

Let us continue the illustrative example in Fig. 5 with 6 measured vectors shown as circular dots. Three black dots represent true patterns, and another three, shown as open dots, are false patterns. We have $\Phi(1,2,1,4,7)=\alpha$ (vertical line 1) because one measured vector related to the true pattern is out of the area described by the system of inequalities (one missed pattern). Similarly $\Phi(1,3,1,4,7)=\alpha+\beta$ (vertical line 2) because there is one false hit and one missing particle, and $\Phi(1,4,1,4,7)=\beta$ (vertical line 3). It is not possible to make the criterion (2) smaller than $\min\{\alpha, \beta\}$ by varying parameters $P_1,\ldots,P_5$ value only. However, by defining $F_5 = P_6V_1 + P_7V_2 - P_5$ and selecting $P_5=6$, $P_6=P_7=1$, the inequality $F_5\leq 0$ describes the area on the left of line 4 in Figure 5 and $\Phi\left(P_1,\ldots,P_7\right)=0$ because the shaded area separates true patterns from the false ones.

Unfortunately, criterion (16) is not convex and the optimization problem is not likely to be solved by standard optimization methods. Heuristics based on specifics of the training samples as well as specifics of the set of inequalities (15) are required. An algorithm is based on selecting reasonable initial values of parameters and uses greedy approach, i.e., optimizes one parameter at a time.

As training by samples is a continuing process, incremental learning is implemented. In this approach, as a new training sample $\mathbf{V}^{N+1}$ is available in addition to the previous N samples, the incremental greedy algorithm adds the value

$$\Delta_{N+1} = \alpha\chi\left(\mathbf{V}^{N+1}\right)\left[1-\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^{N+1}\right)\right]+\beta\left[1-\chi\left(\mathbf{V}^{N+1}\right)\right]\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^{N+1}\right)$$

to the current value of $\Phi(\mathbf{P})$ and, if this value is positive, optimizes those constrains that can decrease Δ_{N+1} without increasing $\Phi(\mathbf{P})$. For example, if $\mathbf{V}^{N+1}$ is a true pattern, i.e. $\chi(\mathbf{V}^{N+1})=1$, but $\hat{\chi}_{\mathbf{P}}\left(\mathbf{V}^{N+1}\right)=0$, the algorithm tries to decrease only those $F_i\left(\mathbf{V}^{N+1},\mathbf{P}\right)$ in (15) that are greater than zero to the point of minimum of overall criterion (16). Incremental greedy approach is especially simple if the constraint used by greedy algorithm includes only one parameter. In this case the one-dimensional optimization problem can be solved by golden search.

4. INFRARED THERMOGRAPHY FOR ELECTRICAL MICRO-DEFECT LOCATION ON FLAT PANEL DISPLAYS

Infrared thermography used for this application was described and justified in section 2, and the technology of defect location as well as test fixtures – in [Enachescu, M., and Belikov, S. (2005)]. An electrical short in a pixel connecting two lines (X and Y connectors in Fig. 6) increases current through those lines, and consequently elevates the temperatures of the lines. Thus the lines, though not defective, may nevertheless appear in the infrared image. The portion of the image that highlights the lines is termed the *defect artefact* of the short. Fig. 7.a shows three representative defect types: point-type (805), line-type (810), and corner-type (815). A point-type artefact is usually related to a strong short with resistance of the shorted pixel significantly higher than the resistance of connecting conductors, while line and corner types – when these resistances are comparable.

The defect artefacts obscure the location of defects. The defect location should be defined differently for each type: for point-type it may be the centre of the particle (805 in Fig. 7.a); for corner-type – intersection of horizontal and vertical lines (815 in Fig. 7.a). The most difficult task is to precisely locate the defect of line-type.

[Enachescu, M., and Belikov, S. (2009)] describes an image-processing technique to detect and classify defect artefacts by type. This technique includes a morphological analysis and definition of each type by systems of inequalities (15) in the space of parameters, although in the reference this system was not explicitly specified. Fig. 9.a demonstrates recognition of three line-type defect artefacts in the images of the infrared camera with large field of view (left in Fig. 1). This will be further discussed below.

Defect location algorithm of line-type artefact (such as in Fig. 7.b) is based on a set of thresholded binary images. Fig. 7.b includes a pair of line-type artefacts LA1 and LA2. As threshold levels increase, line-type artefacts become smaller and thinner, and may disappear or split into several
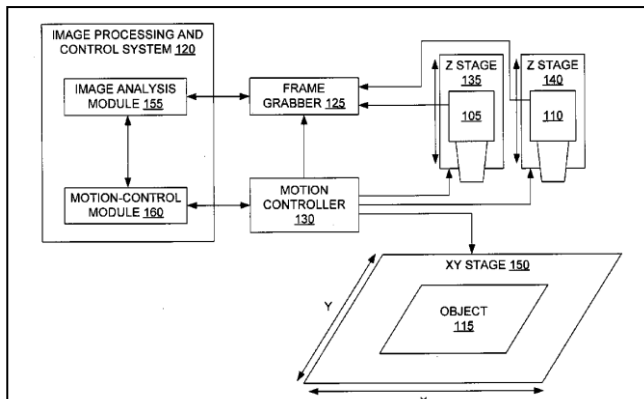
Fig. 6. Inspection system for infrared detection of electrical micro-defects on OBJECT (FPD) [Belikov, S., and Enachescu, M. (2009): Fig. 1]. Photo of two z-stages with infrared cameras is shown in Fig. 1.
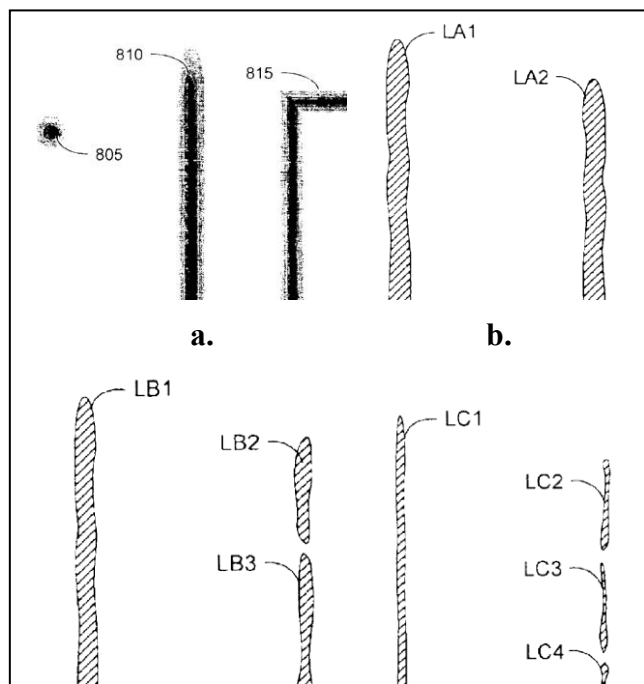

Fig. 7. a. Infrared images exhibiting three types of electrical defects artefacts: point-type (805); line-type (810); and corner-type (815); **b.** Two line-type defects at low threshold; **c.** at intermediate threshold; **d.** at high threshold. [Enachescu, M., and Belikov, S. (2009): Fig. 14, 15].

disconnected lines. In Fig. 7.c, for example, the defect artefact LA2 depicted as a single line in Fig. 7.b appears as a pair of line artefacts LB2 and LB3. The algorithm recognizes that LA2 contains LB2 and LB3. With further increased threshold, we get the image shown in Fig. 7.d with LB2 contains LC2, and LB3 contains LC3 and LC4.

To find the defect location of the line-type defect artefact LA2 (Fig. 7.b) we build the profile such as shown in Fig. 8. In this profile we plot the cross section of the threshold levels vs pixel intensities of the artefacts that belong to LA2, i.e. LA2>LB2>LC2 closest to the edge (LB3 is ignored), and LA1>LB1>LC1. A plot 1725 of Pixel Intensity along a slice
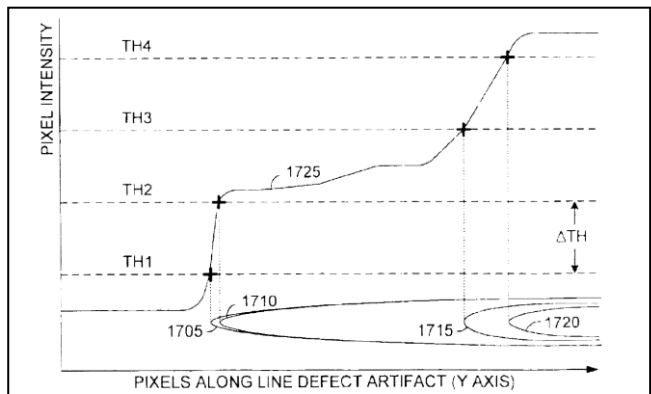
Fig. 8. Illustrative peak profile showing the relationship between four defect artefacts (1705,1710, 1715, and 1720) for different threshold levels. [[Enachescu, M., and Belikov, S. (2009)]]

a. Large field of view –low resolution
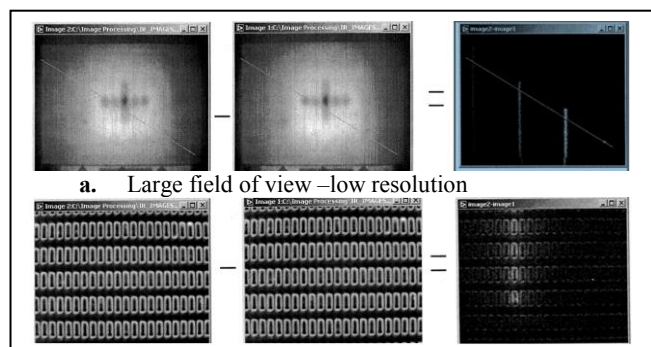
b. Small field of view –high resolution
Fig. 9. Infrared thermography images of inspected FPD

of pixels in parallel with a bisecting the line type defect artefact of the grayscale image illustrates how pixel intensity generally increases as the defect artefact is encountered. Artefacts 1705, 1710, 1715, and 1720 represent slices of plot 1725 taken at four threshold levels. The bold "+" signs indicate the y locations of the peak (i.e. tip) pixels of the respective defect artefacts for different threshold levels.

We define the *defect y-location* as the maximum gradient of the peak profile (Fig. 8). With constant distance ΔTH between the thresholds, the gradient is grad = ΔTH/Δy_i, where y_i are the horizontal locations of the bold "+"s in Fig. 8. A formal algorithm description with related data structures can be found in [Enachescu, M., and Belikov, S. (2009)].

Fig. 9.a demonstrates recognition of three line-type defect artefacts in the images of the infrared camera with a large field of view (left in Fig. 1). Algorithm described above was applied to these images to detect the location of the defect. Then the motion controller (Fig. 6) moves the stage to the position of this location under the high resolution infrared camera (right in Fig. 1). The high resolution infrared images in Fig. 9.b clearly locate the shorted pixel.

5. PARAMETRIC PATTERN RECOGNITION FOR NANOASPERITY DETECTION WITH AFM

In this section we specify vectors **V** of measured variables and **P** of parameters for a practical problem of nanoasperity (NA) detection [Belikov, S., Shi, J., and Su, C. (2008)], [Su,

C., and Belikov, S. (2011)]. Two AFM channels of images are used in this example: height trace and height retrace.

Let $I(y,x)$ and $J(y,x)$ be the (y,x)-pixel values from the first image channel (height trace) and the second image channel (height retrace), respectively.

Let τ be a threshold level for both channels. The measured variables can be divided into two classes: geometrical and regression. Geometrical variables measure geometrical characteristics of the particles on thresholded images. Examples of geometrical variables defined for the particles of both channels are presented in Table 4.

Table 4. Definition of geometrical variables

V_i	Definition	min	max
V_1	Heywood circularity factor $P/(2\sqrt{\pi A})$, where P is perimeter of the particle and A is its area	1	10
V_2	Area A of the particle	200nm^2	∞
V_3	Orientation	-45°	+45°

Parameters "min" and "max" have been optimized using training samples. They are used to select the particles. They applied to both image channels (height trace I and height retrace J).

Regression variables measure certain regression characteristics between different lines of the images in the vicinity of the particles. Let P_I and P_J be related particles selected on the thresholded images I and J; $\left(y^*_{P_I}, x^*_{P_I}\right)$ and $\left(y^*_{P_J}, x^*_{P_J}\right)$ be the locations of the particles, i.e. by definition,

$$I\left(y^*_{P_I}, x^*_{P_I}\right) = \max_{(y,x) \in P_I} I(y,x) \qquad (17)$$

The same definition is true for $\left(y^*_{P_J}, x^*_{P_J}\right)$.

Fig. 10.a demonstrates a feature of a true NA that distinguishes it from false NA. To formalize this feature let us define $\left(y'_{P_I}, x'_{P_I}\right)$ as the largest peak location in the neighbouring lines, i.e.

$$I\left(y'_{P_I}, x'_{P_I}\right) = \max \left\{ I(y,x) \middle| y \in \left\{ y^*_{P_I} - 1, y^*_{P_I} + 1 \right\} \cap P_I, \right\} \qquad (18)$$

and $\left(y'_{P_J}, x'_{P_J}\right)$ is defined by equivalent expression. Particle's pixel values are higher than the threshold level at least for one image I or J. Although $I\left(y'_{P_I}, x'_{P_I}\right)$ or $I\left(y'_{P_J}, x'_{P_J}\right)$ may be lower than the threshold τ, they should be higher than the threshold multiplied by a minimal factor that is a parameter of the optimization problem. Variable V_4 in Table 5 reflects his rule.

Let a_{d,P_I} be the slope of linear regression of the neighbouring lines, i.e.

$$\left(a_{d,P_I}, b_{P_I}\right) = \arg\min_{a,b} \sum_{i=-i_d}^{i_d} \left[I\left(y'_{P_I}, x'_{P_I} + i\right) - aI\left(y^*_{P_I}, x^*_{P_I} + i\right) - b \right]^2 \qquad (19)$$

where i_d is the parameter of the optimization problem (subscript d is for "direct" as opposite to "reverse" as will be described below).
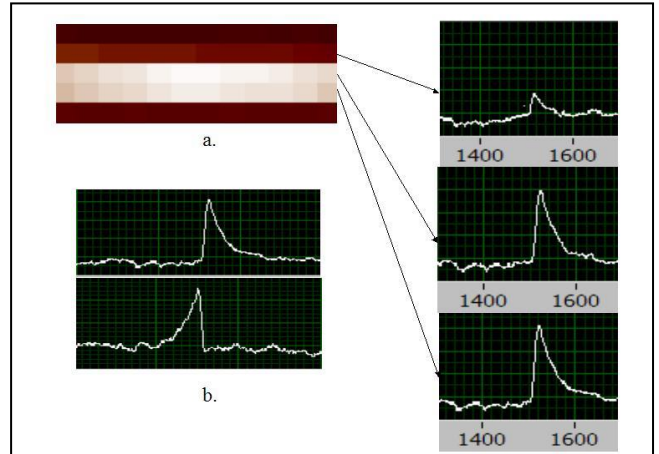
Fig. 10. a. At least one of neighbouring lines strongly correlates with the line that contains NA location in the vicinity of the location; **b.** Comparison of trace (above) and retrace (below) scans of typical nano-asperity. The peak of retrace is on the left of the peak of trace; and one scan strongly correlates with another if reversed.

The variable V_5 of Table 5 reflects the rule that the optimal slopes a_{d,P_I} and a_{d,P_J} should be large enough: a_{d,P_I} at least larger than a specified fraction of $I\left(y'_{P_I}, x'_{P_I}\right) / I\left(y^*_{P_I}, x^*_{P_I}\right)$, and similar for a_{d,P_J}. The specified fraction is a parameter of the optimization problem.

The variable V_6 reflects the rule that regression coefficient

$$R_{d,P_I} = \frac{\mathrm{cov}\left[I\left(y'_{P_I}, x'_{P_I} + \bullet\right), I\left(y^*_{P_I}, x^*_{P_I} + \bullet\right) \right]}{\sigma\left[I\left(y'_{P_I}, x'_{P_I} + \bullet\right)\right] \cdot \sigma\left[I\left(y^*_{P_I}, x^*_{P_I} + \bullet\right)\right]}, \qquad (20)$$
$$\text{where } \bullet = \left\{ -i_d, \dots, i_d \right\}$$

(as well as R_{d,P_J}) should be large enough

Variables $V_7 \geq 0$ (see below) and V_8 are used for rejection related particles on images I and J that are not close enough.

Fig. 10.b demonstrates two other features of true NA that can be used to filter out the false ones. The minimum value of variable V_7 reflects the first rule that location of the NA on image I (trace) must be on the right of the location on image J (this caused by feedback of the scans in opposite directions: trace and retrace). The second feature is the "reverse" correlation of the scans as shown in Fig. 10.b. Similar to (19), (20), let a_{r,P_I,P_J} be the slope of linear regression between the lines on trace and retrace images where NA candidate is located, i.e.

$$\left(a_{d,P_I,P_J}, b_{P_I,P_J}\right) = \arg\min_{a,b} \sum_{i=-i_r}^{i_r} \left[J\left(y^*_{P_J}, x^*_{P_J} - i\right) - aI\left(y^*_{P_I}, x^*_{P_I} + i\right) - b \right]^2 \qquad (21)$$

and equal to V_9 that must be large enough. Similar, "reverse" regression coefficient

$$R_{r,P_I,P_J} = \frac{\mathrm{cov}\left[J\left(y^*_{P_J}, x^*_{P_J} - \bullet\right), I\left(y^*_{P_I}, x^*_{P_I} + \bullet\right) \right]}{\sigma\left[J\left(y^*_{P_J}, x^*_{P_J} - \bullet\right)\right] \cdot \sigma\left[I\left(y^*_{P_I}, x^*_{P_I} + \bullet\right)\right]}, \qquad (22)$$
$$\text{where } \bullet = \left\{ -i_r, \dots, i_r \right\}$$

is equal to V_{10} that also must be large enough. i_r is usually equal to i_d and indicates the number of points at either side of the peaks to use for regression calculations in the formulas.

Table 5. Definition of regression variables for NA detection problem

V_i	Definition	min	max
V_4	$\min\left\{I\left(y'_{P_1},x'_{P_1}\right),J\left(y'_{P_J},x'_{P_J}\right)\right\}/\tau$	0.5	∞
V_5	$\min\left\{a_{d,P_1}\cdot I\left(y^*_{P_1},x^*_{P_1}\right)/I\left(y'_{P_1},x'_{P_1}\right),a_{d,P_J}\cdot J\left(y^*_{P_J},x^*_{P_J}\right)/J\left(y'_{P_J},x'_{P_J}\right)\right\}$	0.6	**1**
V_6	$\min\left\{R_{d,P_1},R_{d,P_J}\right\}$	0.6	**1**
V_7	$y^*_{P_1}-y^*_{P_J}$	0	60nm
V_8	$\left|x^*_{P_1}-x^*_{P_J}\right|$	**0**	750nm
V_9	a_{r,P_1,P_J}	0.3	**1**
V_{10}	R_{r,P_1,P_J}	0.55	**1**

Bold numbers are constants rather than parameters for optimization); the values of parameters are optimized by using training samples.

The nano-asperity experiments used a Dimension™ 3100 AFM. The challenge is to detect the nano-asperity without fault within the data with low signal to noise ratio. The parametric pattern measurement from multiple channels simultaneously provides a robust mechanism in catching the asperities. Figure 11 is a demonstration of the process. Figure 11.a is a survey scan with combined channels for analysis. The circles mark all the qualified patterns with descending asperity height with ascending numbers in the green circles. The survey frame (Figure 11.a) is analysed during real-time imaging and the data are sorted line-by-line. At the end of the frame each qualified pattern will cause the system to automatically perform a high resolution imaging, where the asperity dimension can be precisely measured (Figures 11.b-e). One can see in this example that the nano-asperity cannot be matched with any particular template. However, the parametric pattern detection works successfully for this application.

Fig. 11. (a) Bottom left: the 7 Hz survey scan data analyzed in real time where the true nano-asperities are identified by the circles. (**b**)-(**e**) Zoom in scan at 10 Hz with image resolution of 2 nm/pixel.
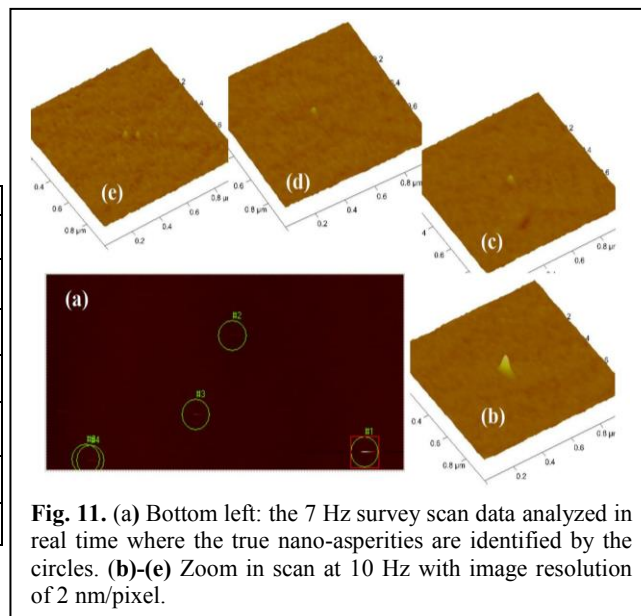
6. CONCLUSIONS

Image-based pattern recognition was demonstrated in two applications: electrical short micro-defects in FPD, and nano-asperity detection on hard disk drive substrates. MEMS mechanical defects were also demonstrated: frictions of the moving parts are higher than nominal friction.

REFERENCES

Belikov, S., et al. (1995). Wafer temperature measurements correction for multi-wavelength imaging pyrometer using Kalman filtering. *Third International Rapid Thermal Processing Conference*, pp. 269-278.

Belikov, S., Shi, J., and Su, C. (2008). AFM based pattern detection for adaptive drift compensation and positioning at the nanometer scale, *2008 American Control Conference*, pp. 2046-2051.

Belikov, S., and Enachescu, M. (2009). Autofocus system and method using focus measure gradient. *US Patent 7,538,815 B1*.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern Classification*. Wiley&Sons.

Gonzalo, L., Paredes, L., and Mullan, J. (2000). Nonlinear filtering for image analysis and enhancement, *Handbook of Image and Video Processing (ed. Al Bovik)*, Academic Press.

Enachescu, M., and Belikov, S. (2005). Methods and systems employing infrared thermography for defect detection and analysis. *US Patent 6,840,666 B2*.

Enachescu, M., and Belikov, S. (2009). Data Structures and Algorithms for Precise Defect Location by Analyzing Artifacts. *US Patent 7,551,769*.

Engineering ToolBox, (2003). *Convective Heat Transfer*. : https://www.engineeringtoolbox.com/convective-heat-transfer-d_430.html

Sharpe, N. Jr. (2002). Mechanical Properties of MEMS Materials. *MEMS Handbook (edited by M. Gad-el-Hak)*, CRC Press, pp. 3-1-3-33.

Muhlstein, C.L., Brown, S.B., and Ritchie, R.O. (2001). High-Cycle Fatigue and Durability of Polycristalline Silicon Thin Films in Ambient Air, *Sensors and Actustors A* 94, pp. 177-188.

Sparrow, E.M., and Cess, R.D. (1978). *Radiation Heat Transfer –Augmented Edition*. Hemisphere Publ. Corporation.

Su, C., and Belikov, S. (2011). Method and apparatus for automatic scanning probe imaging. *US Patent 7,865,966 B2*.

Thomas, L.C. (1993). *Heat Transfer –Professional Version*. PTP Prentice Hall.

Timans, P.J. (1995). The thermal radiative properties of semiconductors. *Advances in Rapid Thermal and Integrated Processing (ed. F. Roozeboom)*, ch. 2, Kluwer Academic Publishing, pp. 35-101