

# Detection of Micro Contamination in Hard Disk Drives Using Angle Measurements and Bayesian Classification



Jirarat Ieamsaard<sup>1</sup>, Frode Eika Sandnes<sup>2,3</sup>, Paisarn Muneesawang<sup>1\*</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, Thailand  
jirirati55@email.nu.ac.th, paisarnmu@nu.ac.th

<sup>2</sup> Institute of information Technology, Faculty of Technology, Art and Design, Oslo and Akershus University College of Applied Sciences, P.O. Box 4 St. Olavs plass, 0130 Oslo, Norway  
Frode-Eika.Sandnes@hioa.no

<sup>3</sup> Westerdals Oslo School of Art, Communication and Technology, Schweigaardsgate 14, 0185 Oslo, Norway

Received 22 July 2017; Revised 22 December 2017; Accepted 8 January 2018

**Abstract.** Micro Contamination has been a major critical defects peculiar to the hard disk drive assembly particularly to head gimbal assembly compartment and it occurs mainly on the air-bearing surface. Consequently, reliable measures to improve the read/write process in a defect of this nature requires an inspection process to check and verify the hard disk components. Therefore, we proposed an image-based automatic inspection method for micro-contamination detection. This proposed inspection approach leverage on angle measurements and Bayesian classification for identifying detects. However, a comparative analysis between this technique and previous angle measurement and intensity threshold methods was carried out to ascertain the performance improvement. The experimental results showed that this method is robust and outperformed the existing methods in literature. Meanwhile, the improvement recorded showed that this method offers a minimal false detection rate for real-time practical applications.

**Keywords:** angle measurement, Bayesian classification, hard disk drive, micro contamination, visual inspection

## 1 Introduction

Hard disk drives (HDDs) are storage device used in virtually every electronics device especially computer devices. Its demand has been on increase due to insatiable storage needs of individuals and cooperate organizations. Recently, users are opting for information storage in cloud, but the irony is that even the cloud services still rely on physical storage. In recent times, hard disk manufacturers are steadily upgrading both in technology and manufacturing processes in order to meet their clients' demands. However, inspection of HDD components is consequently seen as a major manufacturing phase that requires absolute attention so as to avoid future malfunctioning as a result of defects. In production of HDDs several challenges are normally encountered due to miniaturization of HDDs and their components process that occurs. Based on manufacturing plant discussed in this paper, visual inspection machine was used in replacement of human experts for HGAs inspection. COGNEX image analysis system automatically inspects the images automatically captured by cameras unfortunately the system accuracy is low as a result of high false detection rate. Therefore, in other to produce HDD in mass, a new and innovative technology will be required to consolidate the existing methods and guarantee long-term usage and product reliability. In the recent past, several experimental studies on automatic visual

---

\* Corresponding Author

inspection of hard disk defects such as detection of defects on the HDD media surface using spectral imaging [1-2]. Withayachumnankul et al. [3] deployed a filter kernel to detect the edges of hairline crack defects on the surface of piezoelectric actuators. Yammen et al. [4] explored the inspection of corruptions on pole tips at the end of an air-bearing slider using area-based and contour-based features. One of the important problems in HDD manufacturing is the contamination detection problem which is why efforts is geared toward devising a reliable solution. In another image-based attempt [5] three steps were explored to identify contamination on the Air Bearing Surface (ABS) of HGA. Input image of this attempt was initially preprocessed to segment ABS sub-image from the original top-view image of the HGA followed by the identification of potential contamination areas were using circle detection in up-sampled ABS images. Finally, angle measurements and an intensity threshold was used to classify each of the contamination contenders as either a contamination or not. The method proposed in [5] does not rely on template images. However, the angle measurement and intensity threshold used caused many false detections with the low quality test images. Consequence upon that, this study proposes an improved method aimed to reduce the false detection rate.

In this study, emphasis is on the inspection of the ABS which actually is where the read/write head is located during assembling of HGAs. An innovative method is proposed as improvement measures for micro contamination detection. Section 2 presents the measurement of circle angles that are suspected to be contamination. Then the likelihood estimation using cross-covariance feature and the area-based prior value are introduced and it constitutes a Bayesian formula for classification of ABS images. Section 3 shows the experimental result and analysis. The conclusions are drawn in Section 4.

## 2 Method

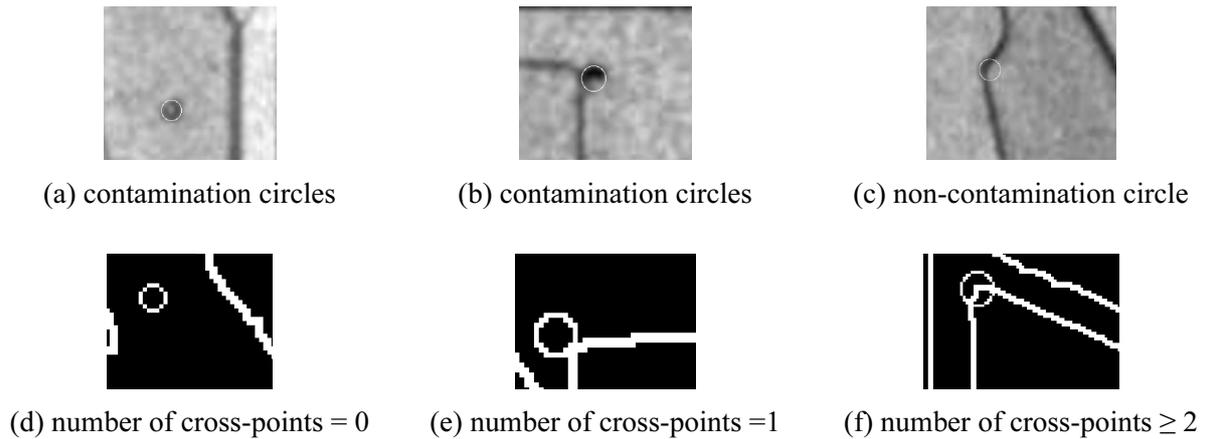
The proposed method starts with a preprocessing step to locate the ABS region in the HGA image. It leveraged on potentialities of circle detection to locate potential contamination areas. However, cross-covariance, likelihood, and angle measurement were used to determine if a potential contamination area is actually contaminated or not. The details of the approach are described in the following sections.

### 2.1 Preprocessing

As mentioned earlier, preprocessing step to locate the potential contamination areas is the first step in realizing the goal of this method. It was achieved using the circle detection and each detected circle is considered based on three cases dependability on the number of cross-point on each circle.

This study used RGB input images of HGAs with a resolution of 2400×2000 pixels (96 dpi) on assumption that the camera capturing the HGA top view was stationary relative to the units on the conveyer belt. A sub-image of the region of interest representing 490×414 pixels was extracted and converted to grayscale. Normalized cross correlation [4, 6] was used to determine the region of interest relative to an ABS template image. The grayscale sub-image is registered with an ABS template image using intensity-based registration [7]. The gray-scale ABS images are up-sampled [8] by a factor of five, since some contaminations are small. Up-sampled images are anti-aliased using a 15×15 median filter [9]. A circle Hough transforms is used to determine potential areas of contamination in the up-sampled image. Circle centers and radii are then transformed back to the co-ordinate system of the original image.

Location of cross-points was done using skeleton of the template. A grayscale version of the ABS template image is low-pass filtered and high pass filtered. The disparity in low pass-filtered image and high pass-filtered image was binarized using Otsu's method [11]. Otsu's method is based on adaptive threshold for image segmentation, whereby the local threshold is computed from the variance within each class. The morphological skeleton operator is subsequently applied to the binarized image of the template image to obtain a skeleton of the ABS binary image. The circles detected in the previous step is now traced to skeleton image with a 50% larger radii ( $radius_{new} = (6/4) \cdot radius_{old}$ ) in order to locate points where the circumference of the detected circles crosses the skeleton [5]. It's imperative to note that each circle is considered as three cases dependent on the number of cross-point on each circle; case:1 the number of cross-points = 0, case:2 the number of cross-points =1, and case:3 the number of cross-points  $\geq 2$ . Fig. 1 and Fig. 2 show the examples of the three cases.



**Fig. 1.** Example of circle detection and circles detected in the skeleton image



**Fig. 2.** Example of detected circles

## 2.2 Angle Measurement

In angle measurement, the two cross-points of A and B and the center C of the detected circle is considered. The line between the cross-points A and C is denoted AC, and that of cross-points B and C is denoted BC. When more than two cross-points existed, more than one angle is calculated and the largest selected [5]. Eq. (1) was used to calculate the angle between two cross-points relative to the circle center. Figure 2 shows examples of angles between two cross points and the circle center.

$$angle = \cos^{-1} \left( \frac{AC \cdot BC}{|AC| |BC|} \right) \quad (1)$$

## 2.3 Cross-covariance Feature

In this work, cross-covariance coefficient (CCC) is used as the similarity scores of two gray scale images for image matching [10] of both grayscale test images and the grayscale template image. This similarity score is used in a subsequent likelihood estimation step. The square region of the grayscale test image and the grayscale template image are detected using the circle centers and radii found during the preprocessing step. The computed cross covariance coefficients of both the test image and template image gives the similarity estimation [10].

$$CCC(A_s, b_s) = 1 - \frac{\sqrt{\sum_{k=1}^{n^2} (q_1[k] - q_2[k])^2}}{\sum_{k=1}^n q_i^2[k]} \quad (2)$$

Here  $\{A_s, [i, k]\}$  denotes the intensity value of the square region of the grayscale template image, and

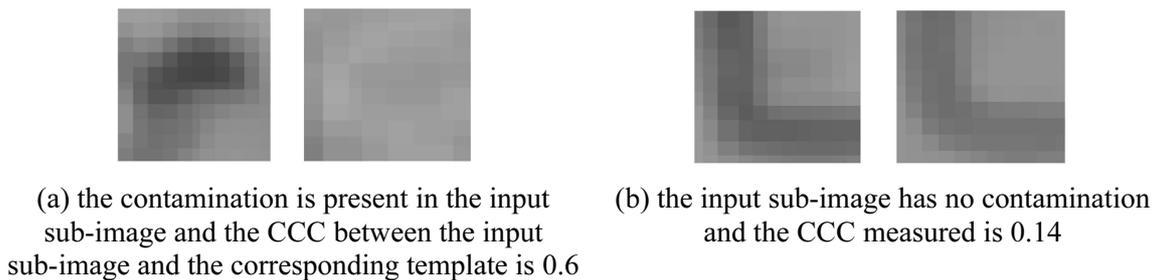
$\{B_s,[i,k]\}$  denotes the intensity value of the square region of the grayscale test image, where  $n$  is the number of pixels. The  $\{q_1[k]\}$  sequence in Eq. (2) is a feature sequence whose elements in descending order are obtained from the auto-covariance matrix of the template image  $\{A_s,[i,k]\}$  that is,  $\{C_A,[i,k]\}$  defined as [10]:

$$C_A[i,k] = \frac{\sum_{x=1}^n \sum_{y=1}^n (A_s[x,y] - \mu_{A_s})(A_s[x-i,y-k] - \mu_{A_s})}{\sigma_{A_s}^2 \times n \times n} \quad (3)$$

Here  $\mu_{A_s}$  and  $\sigma_{A_s}^2$  are the mean and the variance of the template image  $\{A_s,[i,k]\}$ , respectively. Furthermore, the  $\{q_2[k]\}$  sequence in Eq. (2) is a feature sequence whose elements in descending order are obtained from the cross-covariance matrix between two images:  $\{A_s,[i,k]\}$  and  $\{B_s,[i,k]\}$ ; that is  $\{C_{AB},[i,k]\}$  defined as [10]:

$$C_{AB}[i,k] = \frac{\sum_{x=1}^n \sum_{y=1}^n (A_s[x,y] - \mu_{A_s})(B_s[x-i,y-k] - \mu_{B_s})}{\sigma_{A_s} \times \sigma_{B_s} \times n \times n} \quad (4)$$

It may be observed that if the cross-covariance coefficient CCC approaches unity, both images are considered identical. Fig. 3 shows examples of the measurement of the CCC values, where Fig. 3(a) shows the similarity of two images in the case of contamination, and Fig. 3(b) shows the similarity of two images in the case of non-contamination.



**Fig. 3.** The calculation of CCC between two square sub-images

#### 2.4 Likelihood Estimation

Various research efforts have adopted the maximum likelihood function in their work such as [12-14]. In this study, classification of input images was done using a likelihood function of the cross-covariance features. Given  $x$  as the cross-covariance feature vector, and  $P(x|w_i)$  the likelihood of class  $w_i$ , where  $w_1$  is the contamination class and  $w_2$  is the non-contamination class. For the likelihood function:  $P(x|w_1)$  and  $P(x|w_2)$  to be determined in the experiment, cross-covariance features of 427 contamination squares of 313 contaminated ABS images and 23,721 non-contamination squares of 500 uncontaminated ABS images were collected from the training data set. The likelihood function was obtained using histograms [15].

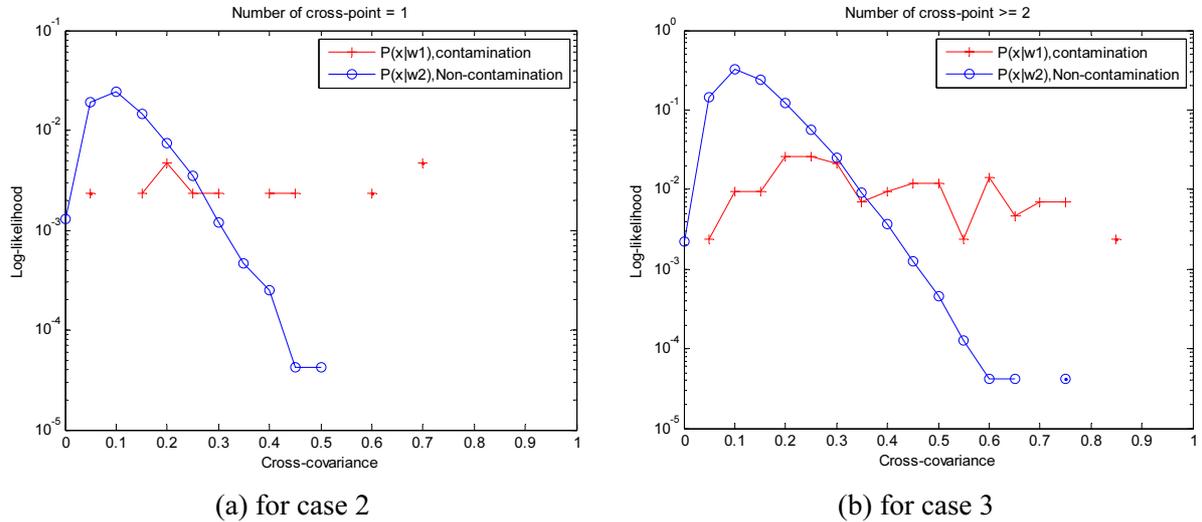
Upon locating the cross-points as shown in Fig. 1, each circle is assigned one of three cases in which a separate likelihood function is considered for each case.

Case 1 – no cross-points: This implies that the detected circle is not present on the ABS template skeleton. Here the probability density of the cross-covariance feature of the training data is used to compute the likelihood of both the contamination and non-contamination classes.

Case 2 – one cross-point: In this case, one pixel of the circumference of the detected circle intersects with the skeleton. The likelihood of both the contamination and non-contamination classes are computed from the cross-covariance of the training data. Fig. 4 (a) shows the log-likelihood  $P(x|w_i)$  of class  $w_i$

(contamination) and the log-likelihood  $P(x|w_2)$  of class  $w_2$  (non-contamination).

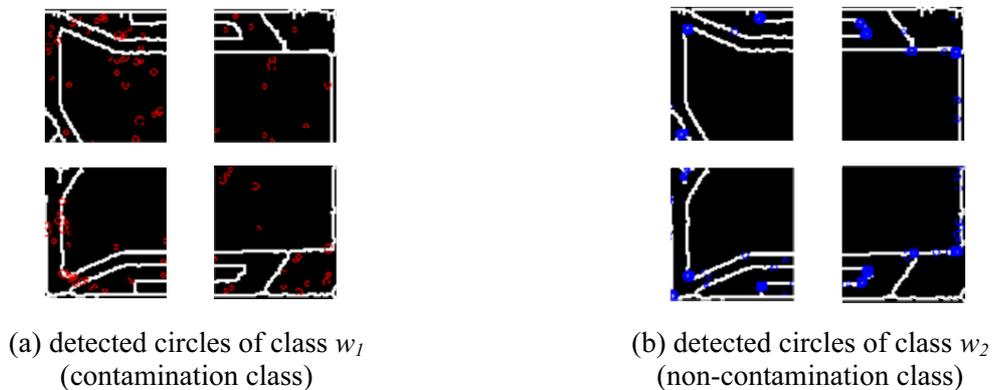
Case 3 – more than one cross-point: This case has at least two pixels of the circumference of the detected circle that intersect with the skeleton of the ABS template. The log-likelihood  $P(x|w_1)$  of class  $w_1$  (contamination) and the log-likelihood  $P(x|w_2)$  of class  $w_2$  (non-contamination) are plotted in Figure 4(b).



**Fig. 4.** Log-likelihood function

### 2.5 Prior Value

Fig. 5 shows the detected circles traced on the skeleton of ABS image from the components of the training data set, which includes 427 contamination circles from 313 ABS images with contaminations and 23,721 non-contamination circles from 500 uncontaminated ABS images. It was found that the non-contamination circles usually occur at the corner of the skeleton and on the rough edge of the skeleton. Therefore, calculation of the prior value was based on this observation in order to classify the detected circle. The prior probabilities of class  $w_1$  and class  $w_2$  are computed from the location of the contamination circles and non-contamination circles detected from the images respectively in the training set. However, skeleton of the ABS template image was subjected to 12 equal sub-regions divisions and the detected circles traced on the skeleton of each region. Then the number of circles that cross the skeleton were counted from each sub-region. The prior probabilities values  $P(w_1)$  and  $P(w_2)$  is computed from the number of circles from class  $w_1$ , and class  $w_2$  respectively.



**Fig. 5.** Example of detected circles which are overlaid with the skeleton in 4 sub-regions

## 2.6 Decision Making

Decision on whether the detected circle represents a contamination or not was done by Bayesian Classifier algorithm. For case 1 and case 2 the likelihood of each case is used to classify the detected circle into one of the two classes, namely the contamination class  $w_1$  or the non-contamination class  $w_2$ . From Bayesian theory one know that  $P(w_i|x)$  varies directly as the  $P(x|w_i)$ .

$$P(w_i | x) \propto P(x | w_i) \quad (5)$$

If  $P(w_1|x)$  is greater than  $P(w_2|x)$ , the detected circle is considered contaminated.

$$x \in w_1 \text{ if } P(x | w_1) > P(x | w_2) \quad (6)$$

To employ the prior probability  $P(w_i)$ ,  $P(w_i|x)$  are obtained from:

$$P(w_i | x) = \frac{P(x | w_i)P(w_i)}{P(x | w_1)P(w_1) + P(x | w_2)P(w_2)} \quad (7)$$

where  $i = [1,2]$  for the contamination and non-contamination classes, respectively.  $P(w_1)$  and  $P(w_2)$  are prior probabilities of class  $w_1$  and class  $w_2$  which are computed from the location of the contamination circles and non-contamination circles (explained in Section 2.5).

For case 3, if the angle is less than a threshold (90 degrees) the circle area is considered a contamination. For circles with angles greater than the threshold, the likelihood and the prior probability according to Eq. (7) are used to make a decision about whether the potential contamination area is contaminated.

## 3 Experimental Evaluation

This proposed method was experimented using 1,363 HGA images. The test images were acquired by a mechanical positioning tool that positioned the camera firmly for capturing. This test suite consists of 1,050 images without contaminations; out of which 500 images were used to calculate the likelihood and the prior probability, and 313 images with contaminations were used to calculate the likelihood and the prior probability. Comparison of the results of this method was done with respect to [5] that used angle measurement and intensity threshold methods to identify contamination in ABS images. The analysis results show when: (1) the proposed method used the likelihood, the prior probability and angle measurement without template registration, (2) the proposed method used only the likelihood without template registration, and (3) the proposed method used the likelihood and angle detection method with template registration. Implementation of this work was done in Matlab and run on a Windows PC and the results shown in Table 1. False detections in non-contamination images are denoted “False Negative”, and false detections in contamination image are denoted “False Positive”.

**Table 1.** Performance comparison of image classification methods

Method	False Negative, %		False Positive, %	
Previous method [5]	187,	17.8%	84,	26.8%
Angle measurement and Bayesian classification without template registration	112,	10.7%	79,	25.2%
Bayesian classification without template registration	228,	21.7%	27,	8.6%
Angle measurement and Bayesian classification with template registration	132,	12.6%	33,	10.5%

The method deployed in [5] yielded a false negative rate of 17.8% and a false positive rate of 26.8%. Hence, this proposed method without template registration yields a false negative rate of 10.7% and a false positive rate of 25.2%. The proposed Bayesian classification algorithm without template registration gives a false negative rate of 21.7 % and a false positive rate of 8.6%. However, by fusing the proposed Bayesian classification and angle detection method with template registration a false negative rate of 12.6% and a false positive rate of 10.5% was achieved. For performance evaluation of this method, the following performance metrics [16] were used, namely F1 score, precision, and recall, with an F1 score reaching its best value at 1 and its worst at 0. Table 2 shows the results.

**Table 2.** Performance evaluation

Method	Precision	Recall	F1 score
Previous method [5]	0.7	0.6	0.6
Angle measurement and Bayesian classification without template registration	0.7	0.7	0.7
Bayesian classification without template registration	0.9	0.6	0.7
Angle measurement and Bayesian classification with template registration	0.9	0.7	0.8

From the experimental results, the combined approach of this proposed method achieved a precision of 0.9, a recall of 0.7, and F1 score of 0.8, against the previous method in literature which recorded a precision of 0.7, a recall of 0.6, and F1 score of 0.6. The algorithm used in this work which deployed Bayesian classification and angle detection method with template registration provides the highest overall F1 score of 0.8. It achieved the precision of 0.9 to equal the precision of the method that used Bayesian classification without template registration, as well achieved the recall of 0.7 to equal the recall of the method that used Angle measurement and Bayesian classification without template registration.

## 4 Conclusions

In this paper, a robust method that fused an angle feature and a Bayesian classifier for detecting micro-contamination on the ABS of the hard disk drive head gimbal assembly (HGA) was proposed. Experimental results showed a reasonable performance improvement with the proposed method when compared with other existing methods.

## Acknowledgements

Sponsorship of this viable project came from the Thailand Research Fund (TRF). This paper is an extended version of [17].

## References

- [1] Z.S. Chow, M. Po-Leen Ooi, Y.C. Kuang, S. Demidenko, Low-cost automatic visual inspection system for media in hard disk drive mass production, in: Proc. of Instrumentation and Measurement Technol Conf (I2MTC), 2012.
- [2] Z.S. Chow, M. Po-Leen Ooi, Y.C. Kuang, S. Demidenko, Automated visual inspection system for mass production of hard disk drive media, Elsevier, Procedia Engineering 41(2012) 450-457.
- [3] W. Withayachumnankul, P. Kunakornvong, C. Asavathongkul, P. Sooraksa, Rapid detection of hairline cracks on the surface of piezoelectric ceramics, Int. J. Adv. Manuf. Technol. 64(9-12)(2013) 1275-1283.
- [4] S. Yammen, P. Muneesawang, An advanced vision system for the automatic inspection of corrosions on Pole Tips in hard disk drives, in: Proc. IEEE Trans. Compon. Packag. Manuf. Technol., 2014.
- [5] J. Ieamsaard, P. Muneesawang, F.E. Sandnes. Image based contamination detection on hard disk head gimbal assembly, in: Proc. Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2015.
- [6] P. Liang, X. Zhiwei, D. Jiguang, Fast normalized cross-correlation image matching based on multiscale edge information, in: Proc. International Conference on Computer Application and System Modeling (ICCASM), 2010.
- [7] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P. Pluim. Elastix: a toolbox for intensity-based medical image registration, Medical Imaging, IEEE Transactions on 29(1)(2010) 196-205.
- [8] R. Fattal, Image upsampling via imposed edge statistics, ACM Transactions on Graphics (TOG) 26(3)(2007) 95.
- [9] T. Nodes, G. Neal, Median filters: some modifications and their properties, in: Proc. IEEE Transactions on Acoustics Speech and Signal Processing, 1982.

- [10] P. Muneesawang, N. Zhang, L. Guan, *Multimedia Database Retrieval*, Springer International Publishing, Geneva, Switzerland, 2014.
- [11] N. Otsu, A threshold selection method from gray-level histograms, *Automatica* 11(285-296)(1975) 23-27.
- [12] D. Barker, Seeing the wood for the trees: philosophical aspects of classical, Bayesian and likelihood approaches in statistical inference and some implications for phylogenetic analysis, *Biology & Philosophy* 30(4)(2015) 505-525.
- [13] B.U. Park, S. Léopold, Z. Valentin, Categorical data in local maximum likelihood: theory and applications to productivity analysis, *Journal of Productivity Analysis* 43(2)(2015) 199-214.
- [14] P.S. Sisodia, V. Tiwari, A. Kumar, Analysis of supervised maximum likelihood classification for remote sensing image, in: *Proc. Recent Advances and Innovations in Engineering (ICRAIE)*, 2014.
- [15] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, 2012.
- [16] H. Huang, H. Xu, X. Wang, W. Silamu, Maximum F1-score discriminative training criterion for automatic mispronunciation detection, *IEEE/ACM Trans. Audio, Speech, Language Process* 23(4)(2015) 787-797.
- [17] J. Jeamsaard, F.E. Sandnes, P. Muneesawang, Detection of micro contamination in hard disk drives using maximum likelihood estimation and angle detection, in: *Proc. JCSSE*, 2016.