

## Image Enhancement Background for High Damage Malay Manuscripts using Adaptive Threshold Binarization

Sitti Rachmawati Yahya<sup>#</sup>, Khairuddin Omar<sup>#</sup>, Siti Norul Huda Sheikh Abdullah<sup>#</sup>, Ali Sophian<sup>\*</sup>

*# Center for Artificial Intelligence Technology, Faculty of Information Science and Technology,  
Universiti Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia  
E-mail: sitti.rachma@gmail.com, ko@ukm.edu.my, snhsabdullah@ukm.edu.my*

*\* Department of Mechatronics Engineering, Faculty of Engineering,  
International Islamic University Malaysia, Jalan Gombak, 53100, Kuala Lumpur, Malaysia  
E-mail: ali\_sophian@iium.edu.my*

---

**Abstract**— Jawi Manuscripts handwritten which are kept at Malaysia National Library (MNL), has aged over decades. Regardless of the intensive sustainable process conducted by MNL, these manuscripts are still not maintained in good quality, and neither can easily be read nor better view. Even though, many states of the art methods have developed for image enhancement, none of them can solve extremely bad quality manuscripts. The quality of old Malay Manuscripts can be categorized into three types, namely: the background image is uneven, image effects and image effects expand patch. The aim of this paper is to discuss the methods used to value add the quality of the manuscript. Our proposed methods consist of several main methods, such as: Local Adaptive Equalization, Image Intensity Values, Automatic Threshold PP, and Adaptive Threshold Filtering. This paper is intended to achieve a better view image that geared to ease reading. Error Bit Phase achievement (TKB) has a smaller error value for proposed method (Adaptive Threshold Filtering Process / PAM) namely 0.0316 compared with Otsu's Threshold Method / MNAO, Binary Threshold Value Method / MNAP, and Automatic Local Threshold Value Method / MNATA. The precision achievement (namely on ink bleed images) is using a proposed method more than 95% is compared with the state of the art methods MNAO, MNAP, MNATA and their performances are 75.82%, 90.68%, and 91.2% subsequently. However, this paper's achievement is using a proposed method / PAM, MNAO, MNAP, and MNATA for correspondingly the image of ink bleed case are 45.74%, 54.80%, 53.23% and 46.02%. Conclusion, the proposed method produces a better character shape in comparison to other methods.

**Keywords**— adaptive equalization; automatic threshold; ink bleed; malay manuscript.

---

### I. INTRODUCTION

Image enhancement is part of an investigation in the old manuscript recognition. It is also part of the pattern recognition. There are several applications of old manuscript recognition such as Tamil text manuscripts, old Damascus Manuscripts, Farsi Manuscripts, Greek Manuscripts, Arabic Manuscripts. They have proposed a statistical framework for degraded document binarization images based on the concept of conditional random fields (CRFs) [1]. The CRFs are discriminative graphical model which conditional distribution model and used in structural classifications. They have determined weights to combine eight orientations to characterize a particular pixel by using QR decomposition method is proposed [2]. By using QR decomposition method, computational complexity is low and weights obtained for each orientation are optimal. Al-Qudah *et al* [3] who expounds a simple threshold technique to

eliminate unwanted background of document images captured by handheld devices. T. Celik [4] have novelty of an algorithm proposed, which enhances the contrast of an input image using spatial information of pixels. The algorithm introduces a new method to compute the spatial entropy of pixels using spatial distribution of pixel gray levels. This researcher [5] presents a new framework for the binarization of broken and degraded document images and restoring the quality of the degraded images. Non-local means method is extended and used to remove noises from input document image in the step of pre-processing.

The aim of old Malay Manuscript in image enhancement is to increase the quality images for a look so brighter than original manuscript. In addition, the enhanced images do not contain any noise and to achieve better accuracies. The Jawi Malay manuscripts that located in the State Library of Malaysia (PNM) has been more than thousand years. Thus far, the process of preserve manuscripts carried out by PNM

still no longer able to produce a good quality as well as easy and clear for readings. In general, the preservation process is to use a special chemical. There are seven preservation manuscripts [6]: cleaning, testing acid content, treating, drying, traditionally repair, repairing use leaf casting machines and binding. These old manuscripts usually experience varying degrees of quality. One of the quality levels in the manuscript is that of thin and paper is easily torn. Additionally, characters have written with thick or thin ink, causing the ink to penetrate the back or front of a page.

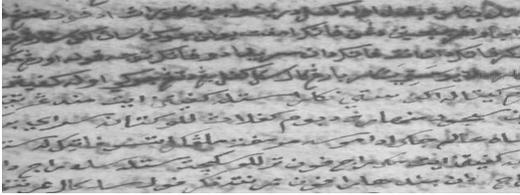


Fig. 1 An example of a Hang Tuah Manuscript image that has an uneven background image quality level.

Researcher [7] propose a straightforward and efficient fusion-based method for enhancing weakly illumination images that uses several nature image processing techniques. Fusion-based framework, images under different weak illumination conditions such as backlighting, non-uniform illumination and nighttime can be enhance.

Other than, there is also [8] have succeed revisit the turbulence recovery problem, incorporating state of the art registration and blur algorithms as building blocks within the sum and blur framework. The performance of such turbulence recovery algorithms has so far been limited. In the last decade, significant progress has been achieved in non-rigid registration and blur image. Many people using local features as a method that producing a better binary document image. As like researchers [9] addressing to the problem of binarization document image with possible stains and general document background information is remove from the image through a background removal stage. The remaining misclassified background and character pixels are separate using a Local Co-occurrence Mapping, local contrast and a two-state Gaussian Mixture Model. Finally, some isolate misclassified components are remove by morphology operator.

The Methods of separating objects from backgrounds can also be implemented using image features. Their research consist the proposed model of three standard steps: preprocessing, main binarization, and post-processing [10]. In the preprocessing and main binarization steps, the features used are mainly phase derived, while in the post-processing step, specialized adaptive Gaussian and median filters are considered. However, object separation in this paper does not involve the high damage documents and printed documents. There are also no researchers applying that technique a novel supervised binarization method is proposed, in which a hierarchical deep supervised network (DSN) architecture is learned for the prediction of the text pixels at different feature levels [11].

The image enhancement technique was introduced by Parker *et al* [12] who have the pre-processing an image include removal of noise, edge or boundary enhancement, automatic edge detection, automatic contrast adjustment and

segmentation. Image enhancement techniques are have suggested also by identifying local pixels to overcome the problem of missing significant pixel information due to the less significant gray level differences [13]. But, this method is not seen for different ink on a particular document that reflects different frequency of light. This paper covers four parts. The first part is the introduction, the second is material and method, in the second about the collection of data, next the third is result and discussion, after that conclusion at the fourth.

Contrast variations on documents are cause usually by many factors, such as noises as page edges, ink leaks, and dirt or any other physical contamination [32]. To reduce the factors that cause excessive contrast, it is necessary to soften the texture of the image. In addition to reducing excessive contrast also to know the pixels that have strong contrast. The local binarization approach displayed good results in uneven illumination cases in which gaps occurred between pixel intensities [33].

## II. MATERIAL AND METHOD

Traditional way of preservation is very complicated and sometimes expensive. Therefore, researchers have offered a variety of image processing methods that are cheaper and without disturbing the original structure of manuscripts. As explained by [14], that some of the researchers succeeded in producing a variety of repairs algorithm, cleaning and seperating texts or objects from the background. One of the methods of image expansion that produced clear images was suggested by [15]. They produce the value of the transformation of an image  $f'(x,y)$  as follows:

$$f'(x,y) \equiv R \times [f(x,y) - a \times m(i,j)] + m(i,j)^b, \quad (1)$$

where  $f(x,y)$  is the original image, and  $R$  is the unit of coefficient of level of obvious difference by using the following equation:

$$R = \left( k \frac{M}{\sigma(i,j)+c} \right). \quad (2)$$

while  $m$  is the average value of the gray level,  $\sigma$  is the standard deviation of pixels in a neighborhood (window) image centered on  $(i, j)$  with a specified region,  $M$  is the average value of the input image, "a, b, c, and k" are variables that have a fixed value of  $0.5 < k < 1.5$ ,  $0.5 < a \in \psi_1$ ,  $b \in \psi_2$ ,  $c \in \psi_3$  with  $\psi_1, \psi_2, \psi_3 \subset \mathbb{R}_+$ .

The result of the Malay Manuscript image for Figure 1 when using Equations (1) and (2) is as Figure 2. The background of this image becomes darker even though its writing (text) is evident. Therefore, the original objective of separating text from background image is not yet achieve if it only uses the equation.

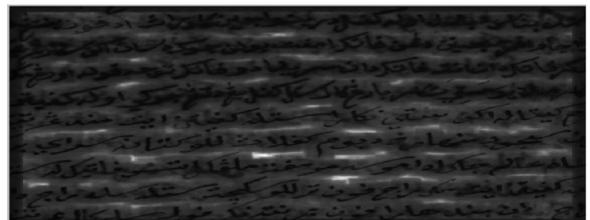


Fig. 2 An example of an old Malay manuscript image tested using equation (1) and equation (2). The background colour of this image is dark so the object is not very clear and the image size is  $640 \times 512$  in 8 bit depth.

Furthermore, [16] suggests a clear stretching method as in the following equation (3),

$$P_k = \frac{(maks-min)}{(f_{maks}-f_{min})} [q_k - f_{min}] + min, \quad (3)$$

with  $P_k$ : colour level of output pixel,  
 $q_k$ : colour level in input pixels,  
 $f_{maks}$ : maximum colour level value in input image,  
 $f_{min}$ : the minimum colour level value in the input image,  
 $maks$  and  $min$ : maximum and minimum colour level values that determine the colour range of the image in an orderly manner,  
 $P_k$ : colour level of input pixel.

Other achievements in the field of analyzing documents and identification of manuscript characters are also proposed by [17]. The character recognition level is doing with an image enhancement method using Laplacian filtering and some mathematical models that successfully capture texts in the image well. This research proposes a framework for overcoming the image quality differences (Figure 1), as described in Figure 3, namely the three stages of the process comprising: Local Adaptive Equalization process (PMT), Image Intensity Value (NII) Process, and Threshold Filtering Process Adj (PAM) in an orderly manner.

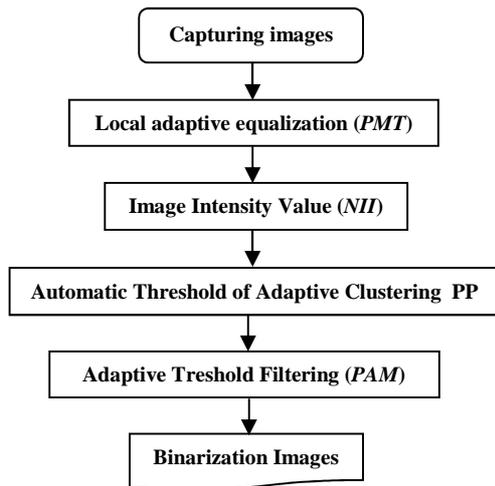


Fig. 3 Proposed a pre-processing framework in text isolation from the old Malay Manuscript image background.

#### A. Local Adaptive Equalization (PMT)

The original image is convert to a gray level of  $640 \times 512$  at a depth of 24 bits. This size has been adopted by DBP because it is a good resolution to store the character information clearly even though the quality level of the manuscript varies. The size of  $640 \times 512$  pixels is more focused on characters in every manuscript page so that it looks more clearly parts of the damage manuscript. To stretch the pixel in an image requires a clear contrast extension to an image. The writer suggests the Local Adaptive Equalization (PMT) algorithm as a refinement of Equation (1) and is defined as Equation (4) below,

$$\hat{g}_{(x,y)} = R \times [g_{(x,y)} - m_{(i,j)}] + m_{(i,j)} + M^2 \times m_{(i,j)}^2, \quad (4)$$

with  $\hat{g}_{(x,y)}$  is the result of the transformation image while  $g_{(x,y)}$  is the input image,  $R$  is the coefficient as equation (2) and fixed value  $k = 0.8$ , grey in a central window  $(i, j)$  and  $M$  is the average of original image.

In this experiment, the author uses window size or region, is  $31 \times 31$  as suggested by [18]. The greater value of the territory then less noise in the background and the darker image object also. Pixels on an object are processed so that the colour of the object is entirely darker using the  $M$  variable which is the average value for 2 dimensions calculated for all the elements in one reference on input image until pixel value on the object has a value close to object's pixel value his neighbours. After that, the background colour is more enlightened to see the difference between the object and the background by using the variable  $m_{(i,j)}$  duplicated. Figure 4 is image of the PMT process.

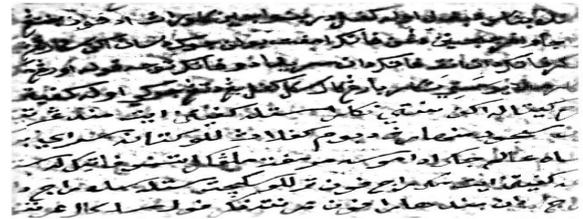


Fig. 4 Result PMT image process with value  $ST = 31$  and value  $k = 0.8$ . Background on the PMT process no darker and brighter than the image in Figure 2.

#### B. Image Intensity Values (NII)

To get the result of an increment from equation (3) for more efficiently, the  $\alpha$  variable not exceeding  $1/3$  is used. To obtain a good result, the equation (5) is change by to add  $\alpha$  parameters:

$$\tilde{g}_{(x,y)} = \alpha \times (\hat{g}_{maks} - \hat{g}_{min}) + \hat{g}_{min} + \hat{g}_{(x,y)} \quad (5)$$

with value of  $\alpha = 0.1$ ,  $max$  and  $min$  is the maximum and minimum pixel value at  $\hat{g}_{(x,y)}$  which is an image result in the PMT process in sequence whereas  $\tilde{g}_{(x,y)}$  is the image of the NII Process. NII can reduce the noise that is still available on the image background to be turn off and become cleaner even though there is little more noise on the image background.

After image is processed using equation (4), the intensity of pixel value will be search by stretching the image histogram. The image of NII process that can be seen in Figure 5 appears to have a less background than the image result in PMT process without reducing black colour of the object.

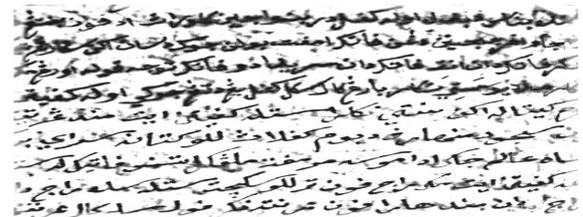


Fig. 5 The results of NII process image appears to have a less background than image results in the PMT process without reducing the black colour of object. NII process layout after adjusting the image contrast with to do stretching of image intensity value at low value is 0.25 and high intensity value image is 0.35.

### C. Automatic Threshold of Adaptive Clustering (PP)

This paper have uses a histogram graph that using the Gaussian window to get a smooth line graph before calculating two of highest peak values. Thereafter, the process to determine background area or foreground of an image on a higher damage image is use of the K-Means clustering technique. Classification of K-Means clustering is determining the position of pixels in two regions, also to help erode some shadows around the text. Then, image process of the damage image is higher to produce a better and clean binary image based on the histogram on two levels. This histogram aims to isolate pixels in the background with the foreground. Algorithm 1 shows the PP algorithm based on two levels of histogram and this process has illustrated in Figure 3 below:

---

#### Algorithm 1 - Automatic Threshold of Adaptive Clustering PP proposed

---

- 1: Start
  - 2:  $f_i$  = Image histogram,  $P_1$  = First highest peak values (background image),  $P_2$  = Second highest peak values (foreground image),  $C_i$  = Clustering on  $i$ ,  $Val\_t_i$  = minimum grey level  $I$ ,  $np_i$  = pixel number in  $i$ , dan  $tmp$  = next step to grey level;
  - 3: Divide the grey level become to 5 class ( $C_i$ ) (each class = 50 pixel number);
  - 4: Find of which group belongs to  $P_1$  and  $P_2$ ;
  - 5: If  $P_1 \geq C_1 \max$   
Then  $r_1 = \frac{C_1 \max}{2}$ ;
  - 6: If  $P_2 \geq r_1$   
Then  $r_2 = \frac{C_2 \max + C_1 \max}{2}$   
 $W_2 = [P_2 : r_1]$   
 $Val\_t_1$  = grey level minimum value on  $W_2$ ;
  - 9: If  $W_3 = [1 : r_1]$   
 $Val\_t_2$  = grey level minimum value on  $W_3$   
Not  
 $W_4 = [C_1 \max : r_2]$   
 $Val\_t_3$  = grey level minimum value on  $W_4$ ;
  - 10: If  $Val\_t_i < 100$   
Then  $tmp_i = tmp - 1$   
 $Val\_np_i(tmp_i) = np_i(tmp_i) - Val\_t_i$   
Not  
 $threshold = Val\_t_i$ ;
  - 11: If  $Val\_np_i(tmp_i) < 100$   
Then  $tmp_i = tmp - 1$   
 $Val\_np_i(tmp_i) = Val\_np_i(tmp_i) - Val\_t_i$   
Not  
 $threshold = Val\_np_i$ ;
  - 12: End
- 

with  $r_1$  and  $r_2$  are the ranges for the first and second batches,  $C_1 \max$  and  $C_2 \max$  are the maximum values of the first and second batch grey level in the method.  $W_2, W_3, W_4$  are the range limits set for  $r_1$  and  $r_2$ ,  $P_1$  and  $P_2$  are the values - the first and second peaks in the corresponding two-stage histogram, and  $f_i$  are the related histogram images. Finally,  $Val\_t, Val\_t_i, Val\_t_2$  is the minimum limit value around to second maximum peak value.

### D. Proposal Filtering Proposition (PAM)

The next process after NII process is the Adaptive Threshold Filtering (PAM) process. In this process, the noise still exists in the object area and background noise on the NII process image ( $\check{g}_{(x,y)}$ ) can be remove using an

appropriate threshold value for the old Malay Manuscript image that used in this writing. In this process, the threshold value used is 100 and the filter used is selectable either average filter or the middle values pixels with intensity range of local neighbours can be made smaller and closer to the mean filter or middle value filter. In this experiment, the author uses an average filter after testing some images. In the image of PAM process, the process NII result is transform into an image binary. The image results after using the NII sub-process then extended to the process using K-Means clustering and binary images using PP automatic threshold values it can be seen in Figure 6.

$$\hat{g}_{(x,y)} = \begin{cases} 1, & \text{if } \check{g}_{(x,y)} < \text{threshold} \\ 0, & \text{if not} \end{cases} \quad (6)$$

with  $\hat{g}_{(x,y)}$  is the output image after binary [19],  $\check{g}_{(x,y)}$  is the result of an image after using the K-Means clustering, and thresh is the PP automatic threshold value based on stage two histogram.

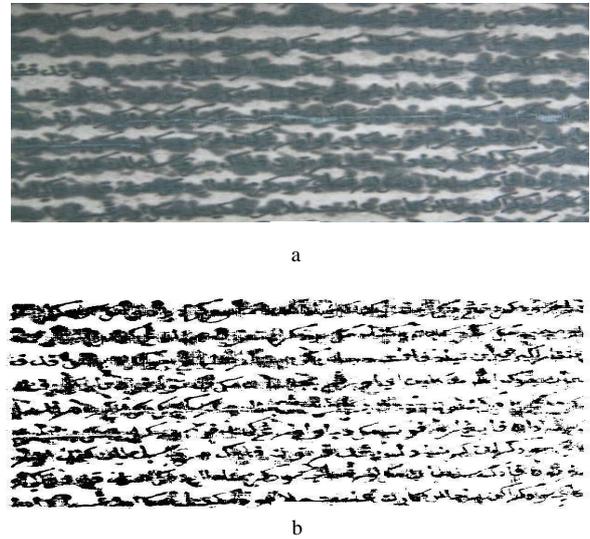


Fig. 6 Image results after NII uses K-Means clustering and PP automatic threshold value into binary images. The MS61 old Jawi manuscript image has a binary threshold  $\check{g}_{(x,y)}$  is 78.

PAM formula can be seen in equations (7) and (8) the following: with  $\check{g}_{(x,y)}$  is the output image after binary [18],  $\check{g}_{(x,y)}$  is K-Means, and  $threshold$  are PP automatic thresholds based on the second level histogram. Image results after using the NII method can be seen on Figure 7.

$$\check{g}_{(x,y)} = (\check{g}_{(x,y)} \times 2) - \gamma, \quad (7)$$

$$\check{g}_{(x,y)} = \delta_{(x,y)} - \check{g}_{(x,y)} - C, \quad (8)$$

$\check{g}_{(x,y)}$  an image as a result of the NII process that has been enhanced by using a threshold value of 100,  $\gamma$  the image of  $\delta_{(x,y)}$  is the average filter process results for a two-dimensional matrix arrangement. At the equation (8),  $\delta_{(x,y)}$  is the result of the NII process that has been using the Median Filtering image for two-dimensional array of matrices, the window size is 20 x 20, and  $C$  is a constant with a value of 0.03. To remove unwanted noise, then try Median filter back with 3 x 3 kernel size to  $\hat{g}_{(x,y)}$  image.

$$\tilde{g}_{(x,y)} = \left[ \bar{g}_{(x,y)} \xleftarrow{\text{median filter}} \check{g}_{(x,y)} \right]^2, \quad (9)$$

with  $\tilde{g}_{(x,y)}$  is a number of images of the Median filter process image in the image  $\check{g}_{(x,y)}$ . Examples of images in Figure 6(b) are still not slick and less clear. Thus, extended using the Median filter and the final result of the image into a binary image can be seen in Figure 7. The image should use a sized filter to reduce excessive noise on the images. After the spots and noise is reduce then the Jawi characters is still in colour. Therefore the formula needs to be repair as below, namely:

$$G = L1 \times L1 \quad (10)$$

with  $G$  is the result image after the binary process,  $L1$  is the result image of the median filter process. The final image result using the Median filter and using the PAM process, noise on image background decreases and Jawi characters on the image start to appear. For PAM's result image can be seen in Figure 7. As for more clearly the result of an example image cleaning phase (using Im61 and Im69 image).

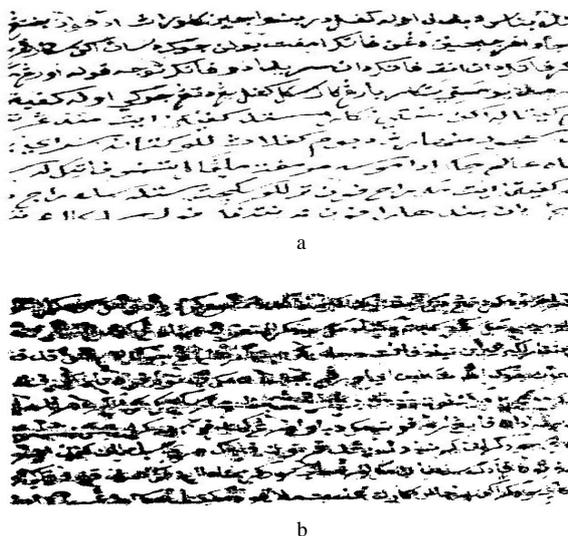


Fig. 7 The result of NII images is processed using equations (6) and (7) and it affects the PAM image. (a) Im69 image, (b) Im61 image.

### E. Image Database

In this experiment, 11 pages of images used that sourced from PNM are divide into three levels of quality, namely: irregular background images, expanded image effects and expanded and blown effect images. The distribution of images into these three levels of quality uses the Signal Method for Noise Ratio (KINH). The result of the eleven part of the image according to the difference in the quality level in this experiment can be seen in Figure 8.

Despite various methods of repairing or cleaning the images are designed [19] [20] [21], [22] mostly solving the problem of manuscripts that have a uniform level of quality. There are still less researchers who conduct research at an uneven or different level of damage. The first group is special for manuscripts with thin ink, then the second group for manuscripts that have patches in certain areas only on

each page, and the third group of manuscripts with very thick ink that almost close every character on each page.

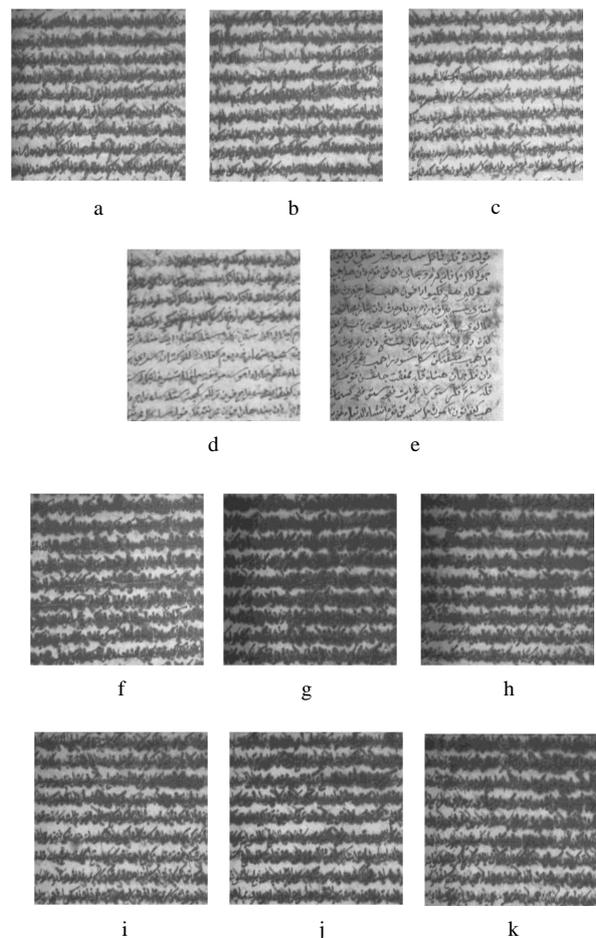


Fig. 8 The 11 pages of Hang Tuah Hikayat's manuscript images used in this paper with varying degrees of quality. Images with irregular background: (a-e). Images that have effect of expanding: (f-h). The image with its impact surface expands as well as impact of streak: (i-k).

The eleven of images are classify according to the quality level of each image. To classify the quality level of The Signal Method for Noise Ratio (KINH) is reserved for image processing. The KINH formula is generally have seen in equation (11), while the signal value and value of the Two Power Source error for noise of RPKD in equations (12) and (13), as follows:

$$KINH = 20 \log_{10} \frac{\text{signal}}{\text{RPKDnoise}}, \quad (11)$$

$$\text{signal} = \mu_{\text{signal}} - \mu_{\text{bkg}}, \quad (12)$$

$$\text{RPKDnoise} = \sqrt{\frac{\sum_{i=1}^n \left( X_i - \frac{\sum_{i=1}^n X_i}{n} \right)^2}{n}}, \quad (13)$$

with  $\mu_{\text{signal}}$  = signal average value,  $\mu_{\text{bkg}}$  = average background value,  $n$  = line numbers in background or region of signal,  $X_i$  = value in row to  $i$  in signal region or background region. Whereas KINH for image processing can be seen in the Table 1 using the formula in equation (14) below which have suggested by [23].

$$KINH = \frac{\mu}{\sigma} \quad (14)$$

with  $\mu$  is the average value of the signal or the expected value and  $\sigma$  is the standard deviation of the estimated noise. In Table 1, three types of quality-based groups are test by their performance using equation (11), equation (12) and equation (13).

To ensure the accuracy level of each image, the signal for the noise ratio (KINH) is more useful for classifying the image according of the quality level that compared to the Bit Error Level (TKBBER) per pixel on each image. Equation (15) is the value of TKBBER used in Table 1 as suggested by [24].

$$TKBBER = \frac{\text{Number of the Bit Error Level (error)}}{N} \quad (15)$$

with  $N$  is the total number of bits.

Equation (15) on the wrong number of bits is useful for knowing most bits per pixel error on each image. In equation (15), to calculate the wrong number of bits (error) is take from equations (16), (17) and (18). Table 1 shows the classification of images based on the TKBBER value and the wrong number of bits on each of the classified images

also. Incorrect number of bit values in Table 1 is sort from minor damage to more damage. Figure 13a-13d below is an example of stages using the proposed method in this experiment using the old Hang Tuah Legendary Manuscript image in Jawi character.

$$error = \sum \frac{signal-r}{2}, \quad (16)$$

$$r = signal + noise, \quad (17)$$

$$noise = noise\_AWGN, \quad (18)$$

with signals is a polar binary data of all the bits ( $N = 10^6$ ). The noise is assumed to be Gaussian and white (AWGN = Averaging of White Gaussian Noise), with a standard deviation of the signal ( $\sigma$ ) derived from equation (19) as suggested by [25] as follows:

$$\sigma_N^2 = 10^{\frac{KINH_{db}}{20}} \quad (19)$$

with  $N$  is the total number of bits, KINH is a signal for a decibel noise ratio.

Figure 9 is an explanation of the value of PAM using KINH accuracy with comparing other methods in this paper.

TABLE I  
CLASSIFICATION OF THE ORIGINAL IMAGE ON 11 OLD IMAGES OF THE JAWI MANUSCRIPT BASED ON KINH AND TKBBER VALUE THAT DIVIDED INTO 3 TYPES OF DAMAGE

	Images with Irregular Background					Images with Its Impact Surface Expands as Well as Impact of Streak			Images that have Effect of Expanding		
	Im63	Im65	Im67	Im69	Im77	Im61	Im109	Im111	Im99	Im101	Im107
<b>Number of Error Bit</b>	13229	20387	12549	8738	9918	17166	18253	20185	18962	23202	20544
<b>KINH</b>	3.7112	3.7395	3.9910	4.4606	3.7112	4.2727	3.4803	3.3796	3.2166	3.3210	2.9682
<b>TKBBER</b>	0.0133	0.0202	0.0125	0.0089	0.0099	0.0172	0.0183	0.0202	0.0189	0.0232	0.0205

All pages of the image in this experiment were compared to the threshold value method studied by some previous researchers. Examples: Otsu method (MNAO) by [26], Binary Threshold Value Method (MNAP) to clean edge of image by [27], The Automatic Local Threshold Value Method (MNATA) by [28]. The performance measurements for each of these methods by the method proposed in this experiment are using the Foreground Area Error Method related (TKB) [29], [30] and Signal Method for Noise Ratio (KINH) using the Average Value Error (RPKB) between the original image (image tested) that already repaired. The results for each image using the TKB method, mean and KINH can be seen in the Table 2, and Table 3 shown in the graph in Figure 9 and Figure 10. For accuracy and recovery values of TKB an image is used the Accuracy Method (Precision) and Recall method described in Table 4 shown in the graph in Figure 11.

The images tested by MNAP are taken from the low threshold value of the method. This is because the image at a low threshold is better for the form of his or her eye even if it is seen directly by the eye. For images tested using MNATA

inequality, the entropy threshold value is determined first then the value is used for the time measurements in the processing used in the method. Table 2 is the TKB values for all tested images compared to the reference image (corrected image) that involves the RPKB value in the KINH formula. Formulas for TKB and KINH referred to below are:

$$TKB = \begin{cases} \frac{A_O - A_T}{A_O}, & \text{if } A_T < A_O \\ \frac{A_T - A_O}{A_T}, & \text{if } A_T \geq A_O \end{cases} \quad (20)$$

with  $A_O$  is the foreground area of the basic truth image and  $A_T$  is the foreground area on the result image of the threshold.

$$sign1 = \sum \left( \frac{\sum f(x,y)^2}{N \times M \times MN} \right), \quad (21)$$

$$KINHs = 10 \log_{10} \left( \frac{sign1}{RPKB} \right) \quad (22)$$

with RPKB is the mean value error between the tested image compared to  $f(x,y)$  which is the image of the reference (image that has been repaired).  $N \times M \times MN$  ( $N$  = row;  $M$  = column) is size of the image window.

TABLE II  
THE VALUE OF TKB ON EACH OF ELEVEN PAGES OF OLD JAWI MANUSCRIPT IMAGES TESTED USING DIFFERENT METHODS.

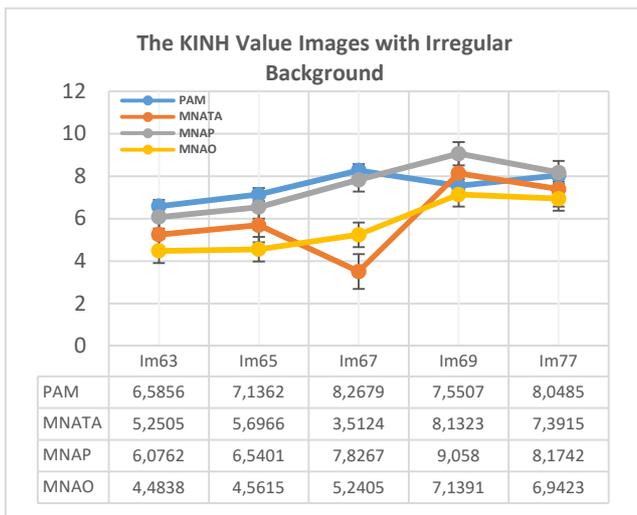
	Images with Irregular Background					Images with Its Impact Surface Expands as Well as Impact of Streak			Images that have Effect of Expanding			Average
	Im63	Im65	Im67	Im69	Im77	Im61	Im109	Im111	Im99	Im101	Im107	
<i>Otsu Method</i> [26] / (MNAO)	200379 .75	198288 .125	223007 .125	269497 .25	286770 .75	156881 .25	129416 .5	157239 .875	189493 .875	171011	3.8567	0.4037
<i>TKB Value</i>	0.5212	0.3885	0.3949	0.3194	0.1248	0.1776	0.6051	0.5201	0.4217	0.4781	0.4890	
<i>Chena et. al. method</i> [27] / (MNAP)	252804 .25	256871 .875	281053 .25	307354 .25	310713 .375	210984 .875	156811 .375	199144 .625	246649 .5	213926	210195 .625	0.2658
<i>TKB Value</i>	0.2285	0.2161	0.1423	0.0620	0.0518	0.3561	0.5214	0.3923	0.2473	0.3471	0.3585	
<i>Ray N. et al. method</i> [28] / (MNATA)	272457 .375	271413 .5	278809 .125	295341 .25	298753 .25	271987 .5	285651 .75	284114	271983 .25	271064 .25	276894 .125	0.1459
<i>TKB Value</i>	0.1685	0.1717	0.1491	0.0987	0.0883	0.1700	0.1283	0.1330	0.1700	0.1728	0.1550	
<i>Proposed method / (PAM)</i>	298658 .125	313454 .875	314156 .125	317325 .75	309936	313430 .5	305329 .5	305922 .875	305459	306786 .625	305696 .875	0.0578
<i>TKB Value</i>	<b>0.0886</b>	<b>0.0434</b>	<b>0.0413</b>	<b>0.0316</b>	<b>0.0542</b>	<b>0.0435</b>	<b>0.0682</b>	<b>0.0664</b>	<b>0.0678</b>	<b>0.0638</b>	<b>0.0671</b>	

TABLE III  
KINH VALUE FROM THE REFERENCE IMAGE OF EACH ELEVEN PAGES OF OLD JAWI MANUSCRIPT IMAGES TESTED ON FOUR DIFFERENT METHODS.

	Images with Irregular Background					Images with Its Impact Surface Expands as Well as Impact of Streak			Images that have Effect of Expanding		
	Im63	Im65	Im67	Im69	Im77	Im61	Im109	Im111	Im99	Im101	Im107
<i>Otsu method</i> [26] / (MNAO)	4.4838	4.5615	5.2405	7.1391	6.9423	3.1176	2.7373	3.3841	4.428	4.3958	3.8567
<i>Chena et. al. method</i> [27] / (MNAP)	6.0762	6.5401	7.8267	9.058	8.1742	5.0754	3.5357	4.5685	5.8457	5.5713	5.0578
<i>Ray N. et al. method / [28] / (MNATA)</i>	5.2505	5.6966	3.5124	8.1323	7.3915	4.7984	3.5721	4.2052	5.2309	4.9619	4.5306
<i>Proposed method / (PAM)</i>	<b>6.5856</b>	<b>7.1362</b>	<b>8.2679</b>	<b>7.5507</b>	<b>8.0485</b>	<b>6.043</b>	<b>4.1398</b>	<b>4.8786</b>	<b>5.9963</b>	<b>5.6855</b>	<b>5.3226</b>

TABLE IV  
PRECISION AND RECALL RAE TKB IN EACH OF THE ELEVEN PIECES OF OLD JAWI MANUSCRIPT IMAGES TESTED IN FOUR DIFFERENT METHODS.

	Images with Irregular Background		Images with Its Impact Surface Expands as Well as Impact of Streak		Images that have Effect of Expanding	
	Precision RAE	Recall RAE	Precision RAE	Recall RAE	Precision RAE	Recall RAE
<i>Otsu method</i> [26] / (MNAO)	75.82%	54.80%	47.97%	20.63%	57.51%	24.56%
<i>Chena et.al. method</i> [27] / (MNAP)	90.68%	53.23%	61.31%	21.42%	73.07%	25.34%
<i>Ray N. et. al. method</i> [28] / (MNATA)	91.2%	46.02%	91.03%	27.34%	89.32%	26.63%
<b>Proposed method / (PAM)</b>	<b>100%</b>	<b>45.74%</b>	<b>100%</b>	<b>27.23%</b>	<b>100%</b>	<b>27.03%</b>



(a)

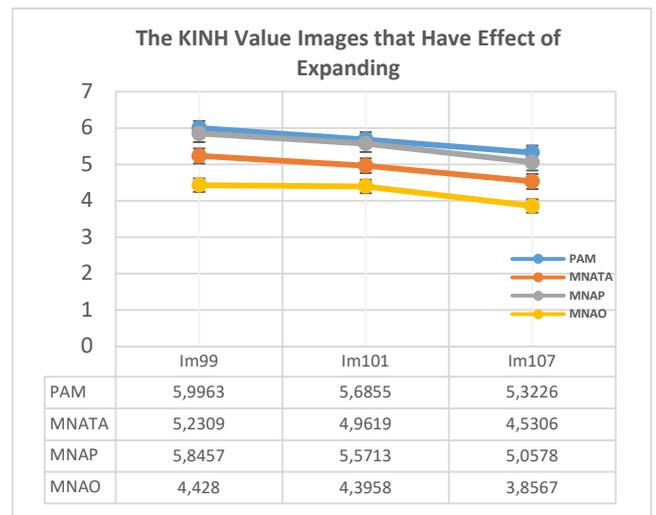
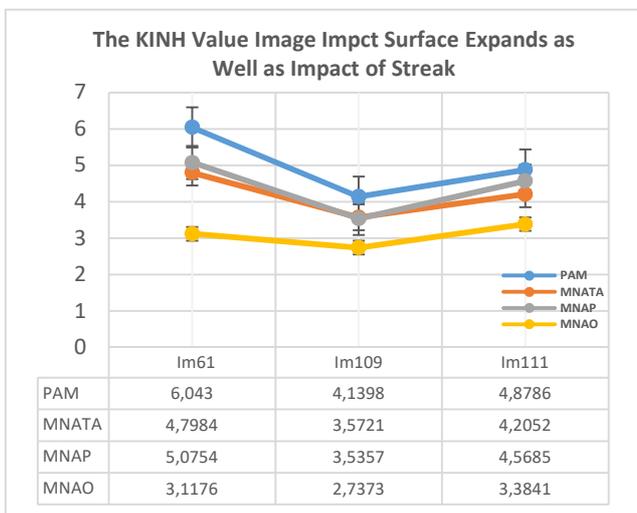
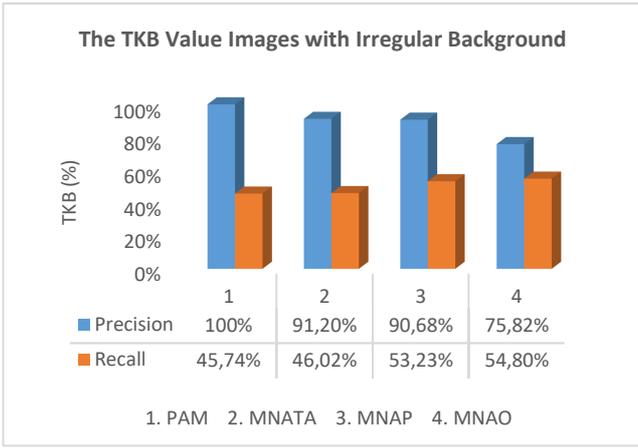


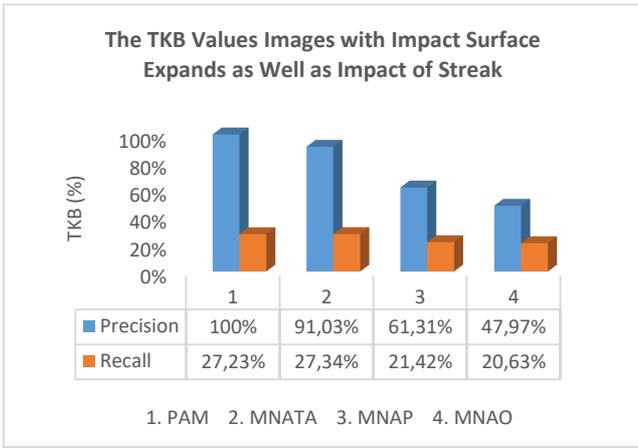
Fig. 9 This figure 9 is refer direct (c) e 3. At (9a) The Im69 image is an image on an uneven background that having a larger KINH value (except PAM Method) than any other image of each method. (9b) The Im99 image is the image on the expanding effect of having a KINH value greater than the other image on each method. (9c) The Im61 image is the image of the expanded effect of the spots having a KINH value greater than the other image in each method.



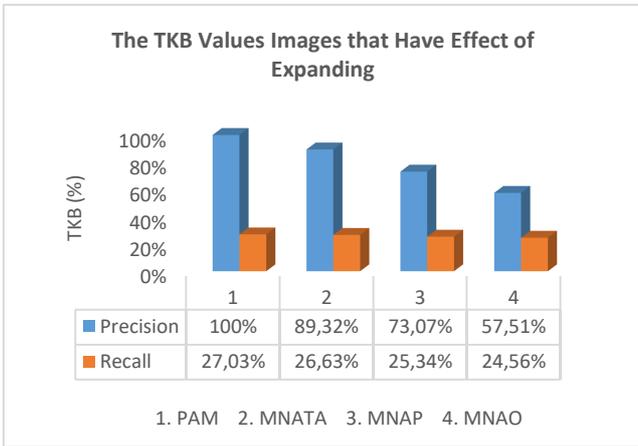
(b)



(a)

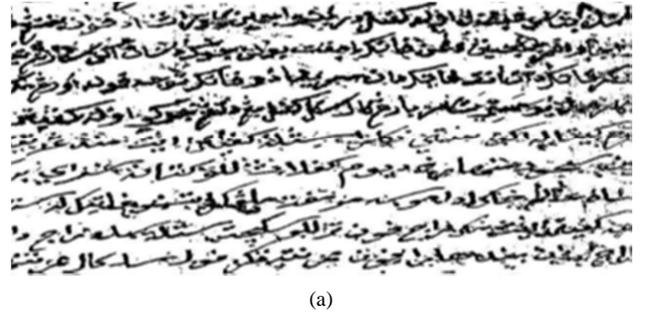


(b)

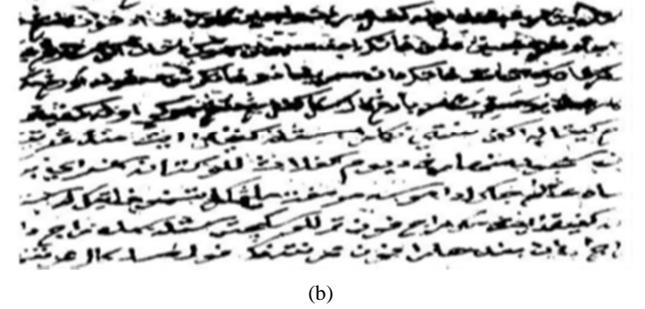


(c)

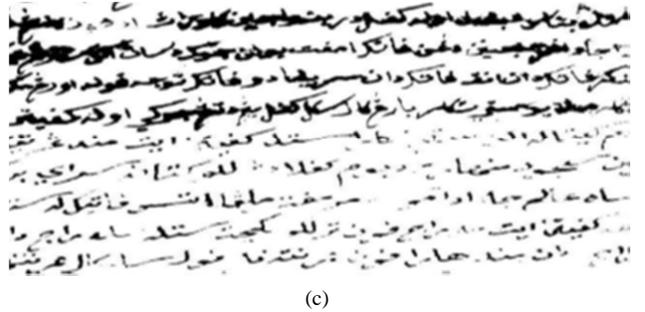
Fig. 10 This figure 10 is refer direct to Table 4. The TKB value of PAM accuracy is greater than the accuracy of the other method on the uneven background image (10a), on the expanding effect image (10b), on the image of the expanded effect of the streak (10c).



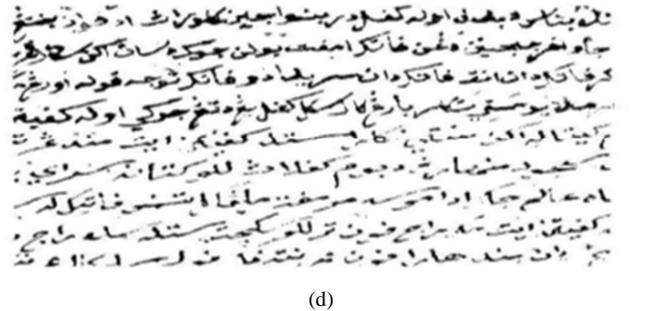
(a)



(b)



(c)



(d)

Fig. 11 Difference of Im69 image results per method used in this experiment. (11a) is the result of Im69 image using MNAO Method, (11b) using MNAP Method, (11c) using MNATA Method and (11d) using Suggested Method (PAM).

### III. RESULT AND DISCUSSION

The results for all experiments are presented in Tables 1 to 4 and are further clarified in Figure 9 to Figure 11. Table 1 shows the value of KINH using the original image. The KINH value in Table 1 is useful for determining the classification of the original image according to the quality level of the original image. Table 2 is shows the value of TKB based on the quality level involving the original image and reference image (the expanded image of the repair). This TKB method evaluates the wrong background area in a reference image. The smaller the TKB value is better image.

The proposed method has a small TKB value compared to the value of TKB on another method so that the average TKB for the proposed method also has a smaller value compared to the value of TKB on another method.

Table 3 is shows the KINH value that involves the original image and the reference image. This is to evaluate the expansion of a reference image. The better and cleaner the background of a reference image is the greater the value of KINH. The difference between TKB Method and KINH Method in Table 2 and Table 3 who is the TKB Method evaluating an image to the character form and the extension of the background improvement, as the KINH Method evaluates an image to the extension of the background repair only. For Table 4, show the two values of RAE value results are that precision and recall. The accuracy value means assessing the perfection of an image while remembering to re-evaluate the accuracy of an image [31].

#### IV. CONCLUSION

Based on the results presented in Tables 1 to 4, it can be concluded that: images that fall into the category of uneven background images have a TKBBER value smaller than that of blistered expanding images and expanding image effects. In Table 1, it can be seen that the Im69 image then the Im77 image is images that have a smaller TKBBER value than the other image TKBBER value. Images in an infinite background image category have a smaller level of quality and with human-level evaluations, uneven background images are still readable and still appear to be clear in the form of their Jawi characters.

The value of KINH in Table 3 and further explained in Figure 9a, Figure 9b and Figure 9c shows that Im69 and Im77 images have a greater KINH value. However, KINH values on Im69 and Im 77 images on PAM are not greater than KINH. Images of Im69 and Im77 on MNAP or Im69 images on MNAP, this does not mean Im69 images on PAM are less favorable as reference images, but instead Im69 images on MNAP or MNATA has a form of underscore characters and a lot of incomplete forms of Jawi characters even though the extension of the background cleaning on Im69 images in MNAP and MNATA is better.

The precision value of the RAE found in Table 4 is higher than the percentage of PAM in every three types of damage which are all worth more than 95%. This precision value is

also clarified in Figure 11 (a, b, c) where the bar chart on the precision of the PAM is higher than the bar graph in other methods. However, for the recall value of PAM which has the highest value compared to other methods, only the image has an expanding effect of 27.03%. Although the percentage value for recalling PAM is not all higher than other methods but does not mean the image on the PAM is poorly improved but instead the reference images on the PAM have the character shape and the result of the improvement is better than the reference images available on the other methods. This is evidenced by the values of TKB values in Table 2.

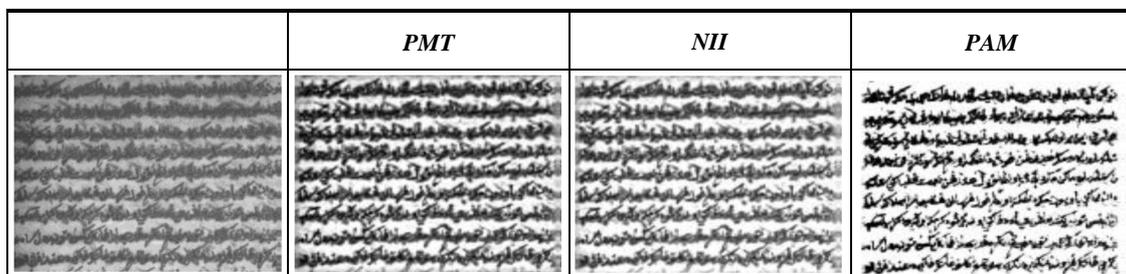
Table 2 shows that the proposed Method (PAM) has a lower TKB value compared to the value of TKB in another method. So, the average value of TKB on PAM is also low. This means that the magnitude of errors in the foreground area of small value reference images on the PAM. For more details can be seen in Figure 10 on the differences of Im69's reference image results in the PAM methodology with the other methods used in the experiment.

The proposed method in this paper shows a good achievement in eliminating background damage in manuscripts with higher damage. In the PAM method that is proposed for the final process of NII is a process that hits the last time the remove of noise are still left on the background of the high damage image. So, that by eliminating the remove of the noise, the shape of the writing characters in the manuscript look more clearly and can still be read. In the Figure 12 also shows all images of the results of the proposed method / PAM with the images of the results of other methods in this paper.

The author realizes in this experiment the number of image data tested is still small, so there must be added more old Malay Manuscript images data at different quality level conditions. So, that the accuracy value for the proposed method is higher and better. Hence, the focus of the author's research is to produce more types of quality levels on the images obtained, such as: high illumination variations, high contrast mix other damage, etc. The standard datasets in this experience are Hang Tuah Manuscripts only that used by MNO, MNAP and MNATA approach.

#### ACKNOWLEDGMENT

This research is partially funded by CAIT under research grant: FRGS/1/2016/ICT02/UKM/01/1.





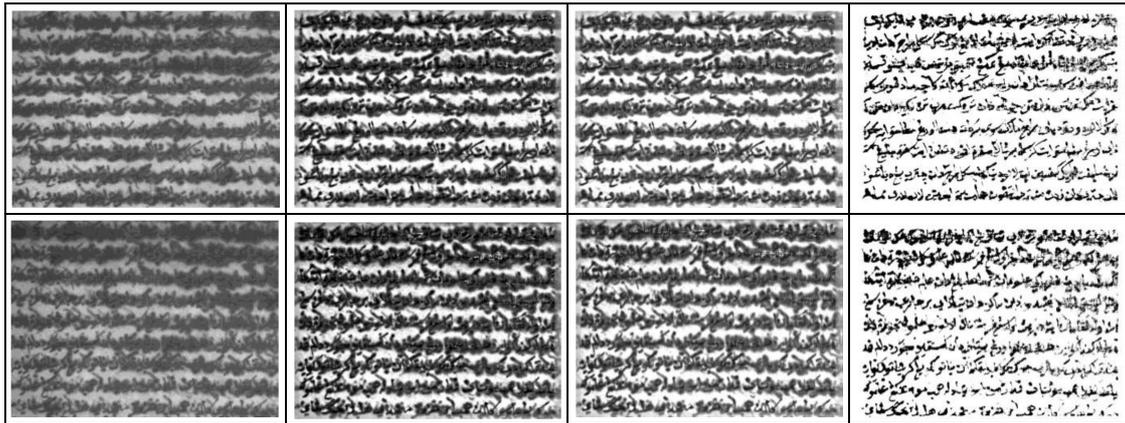


Fig. 12 The differences in Malay manuscript images used in this experiment are classified based on three levels of quality, namely: irregular background image, expanded image effect and expanded-blown effect images through several proposed methods, PMT, NII, PAM. Sorted from top to bottom (images with irregular background: Image Im63, Im65, Im67, Im69, Im77; images with its impact surface expands as well as impact of streak: Image Im61, Im109, Im111; images that have effect of expanding: Image Im99, Im101, Im107).

### REFERENCES

[1] E. Ahmadi, Z. Azimifara, M. Shamsa, M. Famouria, and M. J. Shafiee, "Document image binarization using a discriminative structural classifier", *Pattern Recognition Letters*, vol. 63, pp. 36-42, June. 2015.

[2] F. Kasmin, A. Abdullah, and A.S. Prabuwno, "Weight determination for supervised binarization algorithm based on QR decomposition", *Jurnal Teknologi UTM*, vol. 79, pp. 97-106, January 2017.

[3] Al-Qudah, M. K., M. Nasrudin, B. Bataineh, and K. Omar, "A Novel simple thresholding for uneven illuminated document images captured via handheld devices", *Asian Journal of Information Technology*, vol. 15(16), pp. 2927-2936, 2016.

[4] T. Celik, "Spatial entropy-based global and local image contrast enhancement", *IEEE Transactions on Image Processing*, vol. 23(12), pp. 5298-5308, Dec. 2014.

[5] Y. Chen, and L. Wang, "Broken and degraded document images binarization", *Neurocomputing Journal*, vol. 237, pp. 272-280, May 2017.

[6] Manuscripts, National Library of Malaysia (Perpustakaan Negara Malaysia, PNM) (April 27, 2009), <http://www.pnm.gov.my/pnmv3/index.php?id=84>.

[7] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images", *Signal Processing Journal, Elsevier*, vol. 129, pp. 82-96, December 2016.

[8] R. Gal, N. Kiryati, and N. Sochen, "Progress in the restoration of image sequences degraded by atmospheric turbulence", *Pattern Recognition Letters*, vol. 48, pp. 8-14, Oct. 2014.

[9] N. Mitianoudis, and N. Papamarkos, "Document image binarization using local features and gaussian mixture modelling", *Journal of Image and Vision Computing*, vol. 38(c), June 2015.

[10] H. Z. Nafchi, R. F. Moghaddam, and M. Cheriet, "Phase-Based binarization of ancient document images: model and applications", *IEEE Transactions on Image Processing*, vol. 23(7), pp. 2916-2930, July 2014.

[11] J. Natarajan, and I. Sreedevi, "Enhancement of ancient manuscript images by log based binarization technique", vol. 75, pp. 15-22, May 2017.

[12] J. Parker, O. Frieder, and G. Frieder, "Automatic enhancement and binarization of degraded document images", *Conference: Document Analysis and Recognition (ICDAR)*, August 2013.

[13] S. S. Negi, and Y. S. Bhandari, "A hybrid approach to Image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram", *International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, IEEE, 2014.

[14] S. R. Yahya, S. N. H. Sheikh Abdullah, K. Omar, M. S. Zakaria, and C. Y. Liong, "Review on image enhancement methods of old manuscript with damaged background", *International Journal on Electrical Engineering and Informatics*, vol. 2(1), January 2010.

[15] Munteanu, and A. Rosa, "Gray-scale image enhancement as an automatic process driven by evolution", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34(2), April 2004.

[16] N. Mokhtar, N. H. Harun, M. Y. Mashor, Roseline, N. Mustafa, R. Adollah, Adillah, and N. F. M. Nasir, "Image enhancement techniques using local, global, bright, dark, and partial contrast stretching for acute leukemia Images", *Proceedings of the World Congress on Engineering (WCE)*, vol. 1, London UK, July 2009.

[17] S. Roy, P. Shivakumara, H. A. Jalab, R. W. Ibrahim, U. Pal, and T. Lu, "Fractional poisson enhancement model for text detection and recognition in video frames", *Pattern Recognition*, vol. 52, pp. 433-447, 2016.

[18] M. K. Alqudah, M. F. Nasrudin, B. Bataimeh, M. Alqudah, and A. Alkhatatneh, "Investigation of binarization techniques for unevenly illuminated document images acquired via handheld cameras", *International Conference on Computer, Communications, and Control Technology (I4CT)*, pp. 524-529, 2015.

[19] S. He, P. Samara, J. Burgers, and L. Schomaker, "A Multiple-label guided clustering algorithm for historical document dating and localization", *IEEE Transactions on Image Processing*, vol. 25(11), pp. 5252-5265, Nov. 2016.

[20] J. Wen, S. Li, and J. Sun, "A new binarization method for non-uniform illuminated document image", *Pattern Recognition*, vol. 46(6), pp. 1670-1690, June 2013.

[21] D. N. Satange, S. S. Bobde, and S. D. Chikate, "Historical document preservation using image processing technique", *International Journal of Computer Science and Mobile Computing*, vol. 2(4), pp. 247-255, April 2013.

[22] B. Bataineh, S. N. H. S. Abdullah, and K. Omar, "Adaptive binarization method for degraded document images based on surface contrast variation", *Pattern Analysis and Applications*, vol. 20(3), pp. 639-652, August. 2017.

[23] J. T. Bushberg, J. A. Seiert, E. M. Leidholdt JR, and J. M. Boone, "The essential physics of medical imaging", (2e), *European Journal Of Nuclear Medicine And Molecular Imaging*, Philadelphia: Lippincott Williams & Wilkins, vol. 30(12), pp. 280, 2002.

[24] G. Breed, "Bit error rate: Fundamental concepts and measurement issues", *High Frequency Electronics*, 2003.

- [25] K. A. Phillips, J. H. Reed, and W. H. Tranter, "Minimum BER adaptive filtering", *IEEE International Conference on ICC*, vol. 3, pp. 1675-1680, June 2000.
- [26] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Trans. Sys., Man., Cyber*, vol. 9, Pp. 62-66, 1979.
- [27] Q. Chen, Q. Sun, P. A. Heng, and D. S. Xia, "A double threshold Image Binarization Method Based on Edge Detector", *Pattern Recognition*, vol. 41(4), pp. 1254-1267, 2007.
- [28] N. Ray, and B. N. Saha, "Edge sensitive variational image thresholding", *IEEE International Conference Image Processing*, (ICIP 2007), vol. 6, pp. 37-40, 2008.
- [29] W. Boussellaa, A. Bougacha, A. Zahour, H. El Abed, and A. Alimi, "Enhanced text extraction from arabic degraded document images using EM Algorithm", *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 743-747, 2009.
- [30] C. C. Fung, and R. Chamchong, "A Review of evaluation of optimal binarization technique for character segmentation in historical manuscripts", *Third International Conference on Knowledge Discovery and Data Mining*, pp. 236-240, 2010.
- [31] D. Albashish, S. Sahran, A. Abdullah, N. A. Shukor, and H. S. M. Pauzi, "Lumen-nuclei ensemble machine learning system for diagnosing prostate cancer in histopathology images", *Pertanika J. Sci. & Technol*, vol. 25 (S), pp. 39 – 48, 2017.
- [32] A. Qasem, S. N. H. Sheikh Abdullah, S. Sahran, R. I. Hussain, "An Accurate rejection model for false positive reduction of mass localization in mammogram", *Pertanika J. Sci. & Technol*, vol. 25(S), pp. 49 – 62, 2017.
- [33] A. B. Al-Naqeeb, and M. J. Nordin, "Robustness watermarking authentication using hybridisation DWT-DCT and DWT-SVD", *Pertanika J. Sci. & Technol*, vol. 25(S), pp. 73 – 86, 2017.
- [34] S. M. M. Kahaki, M. J. Nordin, W. Ismail, S. J. Zahra, R. Hassan, "Blood cancer cell classification based on geometric mean transform and dissimilarity metrics", *Pertanika Journal of Science and Technology*, vol. 25(S6), pp. 223-234, June 2017.
- [35] S. Hakak, A. Kamsin, J. Veri, R. Ritonga, T. Herawan, "A Framework for Authentication of Digital Quran", *Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing*, vol. 672, Springer, Singapore, March 2018.