Forgery Detection Based Image Processing Techniques

Shaimaa H. shakir, Nour Zwyer

Abstract: The forgery of official documents becomes familiar and this made a lot of problems and difficulties to the official institutions .One of these problems is the development of digital image processing software and editing tools. with the new the sophisticated powerful digital printers and a lot of software tools it become very simple to edit scanned document and create new one with different information that is very difficult to distinguish from the original and the forgery one. This paper used image processing techniques to detection forgery in official scanned document. The aim of this paper is design a quick and most efficient system for detecting forgery in official documents. the result for detection depend on removing noise if the system applied on scanned document with noise Increases the presence of forging ratio**.** The paper is arranged as follows: Section 2 presents a related works, section 3 introduce classification of image forgery, Section 4 presents the proposed work, Section 5 illustrated the results of system work and Section 6 concludes the paper

Index Terms: forgery, scanned document, image processing techniques, removing noise, detection forgery, official documents

1 Introduction

Nowadays with the new the sophisticated powerful digital printers and a lot of software tools it become very simple to edit scanned document and create new one with different information that is very difficult to distinguish from the original and the forgery one. Documents forgery become widely separated and very common problem all over the world and especially in Iraqi society after 2003 .Many people use this way illegally to get jobs throw out forgery their certificate or even in realestate selling or buying properties . To reduce this phenomenon there are several methods developed to help detect forgery in official documents. Digital image forensics is a field that analyzes images of a particular scenario, in order to establish the credibility and authenticity (or otherwise), through a variety of means. It is fast becoming a popular field because of its potential applications in many domains like intelligence, sports, legal services, news reporting, medical imaging and insuranceclaim investigations[1] [2].

In this paper we tend to present a forgery detection technique that's applicable within the case of pixel based mostly forger. The system work technique depends on each the Discrete-Wavelet -Transform (DWT) and therefore the Principle component Analysis (PCA). The features reduced by first applying DWT to induce the approximate sub-band. This can be followed by dividing the latter to fixed sized square blocks then applying PCA to sub-band (LL) block, so as to reduce the features more your paper.

2 Related works

This part presents different techniques that used for detection forgery in document image in some previous related works. **[Zimba et al 2011]** presents detection forgery cloning image by improved algorithm based on Discrete Wavelet Transform (DWT) reduced the image size to frequency sub-band then perform Principal Component Analysis Eigenvalue decomposition(PCA-EVD) on each row vector to reduce vector length. The performance 100% for the size of duplicated region also affects the detection rate in images with JPEG compression. The larger the region size or the maximal the JPEG quality, or themaximal the Signal to noise ratio [3].

[Lin et al 2011] presents integrated technique for splicing and copy move forgery JPEG compression image using method speeded Up Robust Feature (SURF) to extract feature is applied descriptors for finding copies of the same object. The Disadvantage proposed method focuses on JPEG format only. The performance quick and efficient for detection forgery and its reduced time execution is about 200 second [4].

[Kakar et al 2012] presents post-processed copy paste forgeries using method transform –invariant features these are obtained by using feature from the MPEG-7 image signature tool. The efficacy of this technique in detecting copy–paste forgeries with noise addition, blurring, translation, scaling, flipping, rotation and lossy compression. The performance obtain a feature matching accuracy in excess of 90% across post processing actions and are able to detect the duplicated regions [5].

[Ahmed et al 2014] proposed method for detection forgery in document based on distortion during the forgery creation process. The method used Recognition by Adaptive Subdivision of Transformation space (RAST). In this method, two images are corresponding collected and a matching score is calculated. Improvement in this method up to 29% in accuracy of forgery detection [6].

[Ramzi M. Abed 2015] presents detect forgery in scanned document by divide image into overlapping block then used (GLCM) Gray Level Co-occurrence Matrix and (GLDH) Gray

- *Shaimaa H.Shakir: Computer Science Department, Technology University, Baghdad, Iraq. E-mail: 120011@uotechnology.edu.iq*
- *Nour Zwyer: Iraqi Commission for Computers and Informatics, Informatics Institute for Postgraduate Studies, Baghdad, Iraq. E-mail: nourbasim90@gmail.com*

Level Difference Histogram for extract feature and (LDA+SVM) for classified text document the result over 90% detection accuracy[7].

[Rahmati et al 2016] produce method for detect forgery copy move in digital image used Affine- Scale Invariant Feature Transform(ASIFT).The proposed method detects a large number of matched key points in duplicated regions and final e timates the regions correctly [8].

[Khizar Hayat et al 2017] proposed forgery detection method that depends on the discrete wavelet transform (DWT) anddiscrete cosine transform (DCT) for feature reduction. Divided image to dividable blocks using DWT then apply DCT. The blocks are then compared on the basis of correlation coefficients. The maximum accuracy (94.74%) for the proposed method [9].

3 Classification of Image Forgery

For verification of scanned document several methods have been developed. These methods are generally characterized into two approaches [10]:

1-The active approaches: are mostly concerned with the dta hiding techniques, such as digital watermar ing/copyrighting, the prior information is considered essential and integral to the process.

2-The passive approaches: not require any prior information about the original image. The passive blind techniques, where the analyzer has just the final product, provide an explanation to recognize image changes without relying on the insertion of an extrinsic data or digital signatures for image verification.

A. The copy/move forgery is one of the most well-known type of image forgery whereas a part of the image is copied and pasted on another part at the same image.

B. The image splicing is the same copy-move but here the forger copy or cutting a part of an image and pastes it in the image.

C. The retouching incomes to make some retouch in the image by altering or modification or removing some contents of the

4 Proposed Work

This section explain the proposed work for detection forgery in scanned official document depend on pixel based type. The proposed work consists of three main stages training phase, testing phase, classification). Figure 1 shows the propose work stages.

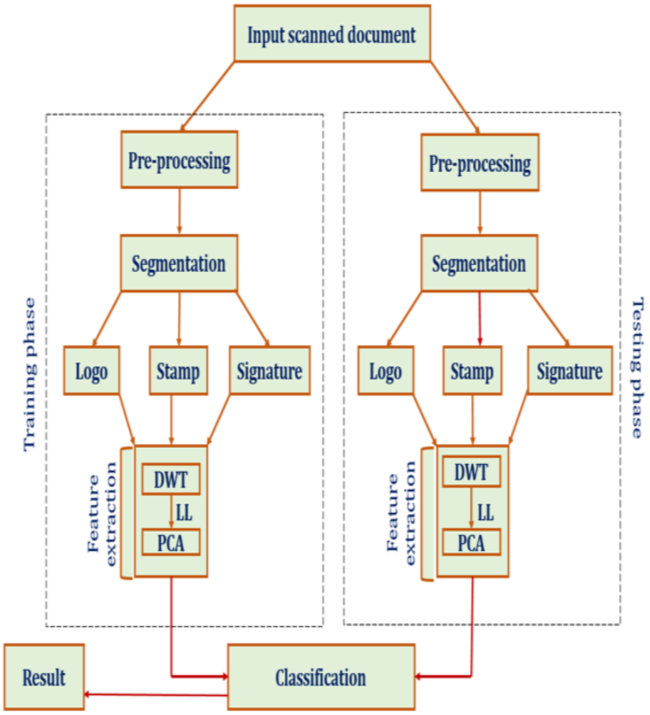


Figure 1 the proposed work

A-Training phase

1-Select scanned document from virtual dataset in order to apply the pre-processing steps-

Step1: convert to gray level if the scanned document is colored.

Step2: Normalization

$$I_N=(I-Min)\frac{NEWMax-NEWMin}{Max-Min}+NEWMin \qquad (1)$$

X$I:\{X \sqsubseteq R^n\} \longrightarrow \{Min,\dots\dots,Max\}$ with intensity values in the range (Min,Max), into a new image with intensity values $I_N:\{X \sqsubseteq R^n\} \longrightarrow \{NewMin,\dots\dots,NewMax\}$in the range (NewMin,NewMax)

Step3: Rotate Correction if the scanned document is suffers from wrap

Step4: Noise Removal using this equation

$$I_{DN} = IDWT(MF(DWT(I))) \qquad (2)$$

Where (MF) mean the median filter function, (I) mean image, (IDWT) mean invers discrete wavelet transfer DWT mean discrete wavelet transfer.

Then save preprocessed image to virtual dataset.

2- Segmented the preprocessed image into three part (logo, stamp and signature) by apply the following steps:

Step1: Read the scanned document from virtual dataset that had been saved in preprocessing step.

Step2: Apply ROI by using function to select the part of (logo, stamp and signature) from scanned document to crop it. Then save the (logo, stamp and signature) images to virtual dataset (file format is JPEG)

3- Feature extraction apply this step:-

Step1: Read the (logo, stamp and signature) images that saved from previous step in virtual dataset

Step2: apply discrete wavelet transform (DWT) on (logo, stamp and signature) images

Step3: Select sub-band LL from DWT

Step4: apply the principle component analysis (PCA) to extract the feature.

B-Testing phase

All the steps that have been made in Training phase are repeated in testing phase.

C- Classification

This step is apply to classify the scanned document as forged or not using minimum distance classifier. Using minimum distance classifier the equation [12]:

$$D(A,B) = \sqrt{\sum_{i=0}^{n} \frac{(|A_i - B_i|)^2}{A_i}} \quad (3)$$

Where D= Euclidean distance between image A and image B, Ai = Feature vector of image A, Bi = Feature vector of image B, n = vector length of vector A and vector B ((sum of textural features).Figure2 illustrate the result end of system work

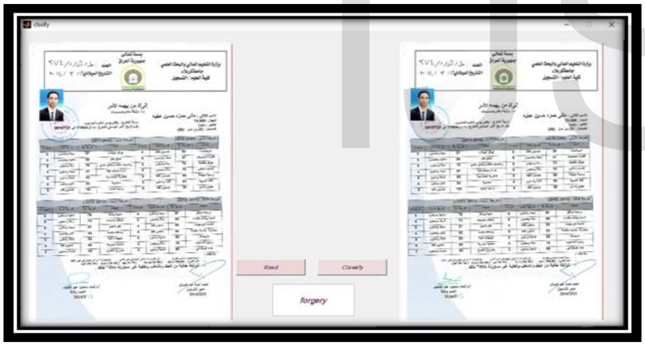


Figure2 the end result of the system work

5 The Result of System Work

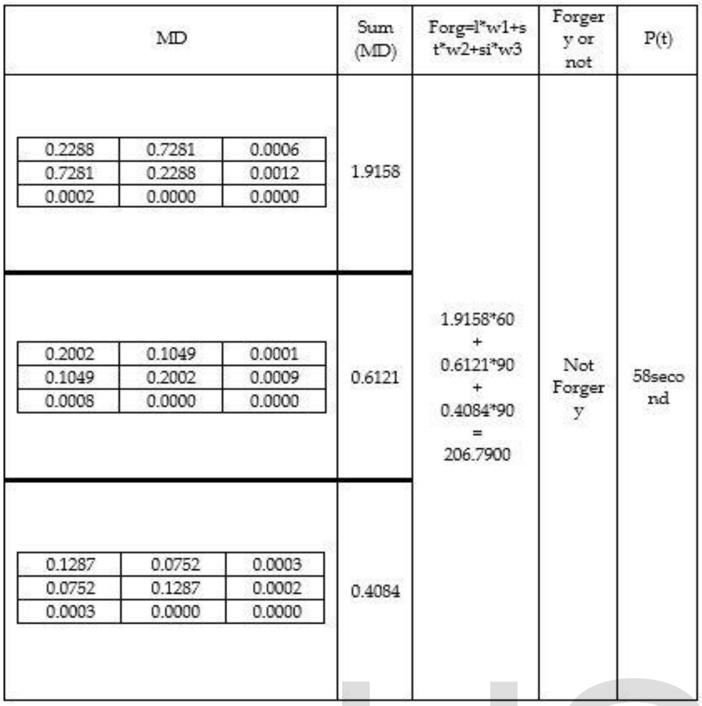
The process of implementing the system and its implementation was done using (Matlab program for programming and design using the graphical user interface) by working on a dataset that had been created personally. The dataset contains 20 original official documents and 20 tampered official documents. The result of system work explained the experiments that applied on proposed work for calculation and discuss results of scanned document. The abbreviations found in the results tables below are: MD mean minimum distance, sum (md) mean summation of all element of matrix for minimum distance and p (t) mean performance of time. The result that determines whether the official document is forged or not depends on this formal: Forg= [L*W1+ST*W2+SI*W3] Where L mean logo, ST mean stamp, SI mean signature and W mean weight. The choice of weight depends on the importance of the part cutting from the scanned document. Stamp andsignature are more important, so weight is given 90 and the logo less important is given weight 60. Since the sum of these weights equals 240, if the resulting value of this formal is less than 240, the scanned document is not considered forged, and if it is larger, the scanned document is forged. The results were extracted in table 1(A, B) after removal of the noise from images. In this table it is clear that there are additions happened to the logo as well as the signature and the stamp contains a little of proportion of modification and change leading to a change in results, this can be seen in the column (sum (MD)). The result of formula smaller than 240. The conclusion is that the document three is not forged. Table2 (A, B) shows the result of scanned document three without removing the noise. In this table the results are different with higher percentages of results in Table 2, which affects the value that will result from the formula and this affects the comparison that determines the document if it is forged or not. In this document concluded that removing the noise is necessary because the presence of noise has an effect, which makes the result appear to be **a forged document.**

TABLE 1(A)

THE RESULT FOR SCANNED DOCUMENT ONE AFTER NOISE REMOVAL

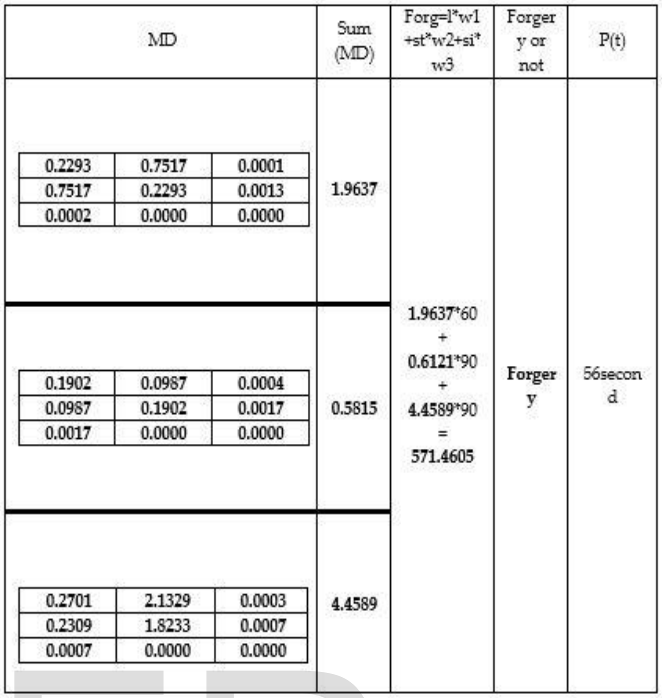
document	part	PCA		
original	logo	0.8377	0.5461	0.0014
		-0.5461	0.8377	-0.0009
		-0.0017	-0.0000	1.0000
forgery	logo	0.9995	0.0312	0.0018
		-0.0312	0.9995	-0.0001
		-0.0018	-0.0000	1.0000
original	stamp	0.3920	0.9200	-0.0011
		0.9200	-0.3920	-0.0025
		0.0027	0.0000	1.0000
forgery	Stamp	0.5336	0.8458	-0.0012
		0.8458	-0.5336	-0.0018
		0.0022	0.0000	1.0000
original	Signature	0.5492	0.8357	0.0006
		0.8357	-0.5492	0.0009
		-0.0010	-0.0000	1.0000
forgery	Signature	0.4582	0.8889	0.0004
		0.8889	-0.4582	0.0007
		-0.0008	-0.0000	1.0000

TABLE 1(B)
SUPPLEMENT TABLE 1(A)



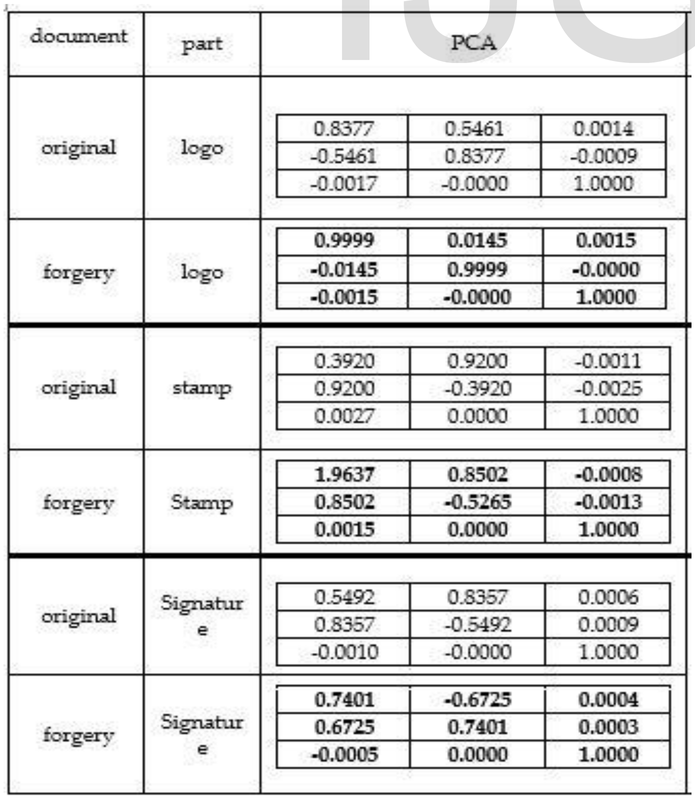
MD	Sum (MD)	Forg=l*w1+st*w2+si*w3	Forgery or not	P(t)
0.2288 0.7281 0.0006 0.7281 0.2288 0.0012 0.0002 0.0000 0.0000	1.9158	1.9158*60 + 0.6121*90 + 0.4084*90 = 206.7900	Not Forgery	58second
0.2002 0.1049 0.0001 0.1049 0.2002 0.0009 0.0008 0.0000 0.0000	0.6121			
0.1287 0.0752 0.0003 0.0752 0.1287 0.0002 0.0003 0.0000 0.0000	0.4084			

TABLE 2(B)
SUPPLEMENT TABLE 2(A)



MD	Sum (MD)	Forg=l*w1+st*w2+si*w3	Forgery or not	P(t)
0.2293 0.7517 0.0001 **0.7517 0.2293 0.0013** **0.0002 0.0000 0.0000**	**1.9637**	**1.9637*60** **+** **0.6121*90** **+** **4.4589*90** **=** **571.4605**	**Forgery**	56second
0.1902 0.0987 0.0004 **0.0987 0.1902 0.0017** **0.0017 0.0000 0.0000**	**0.5815**			
0.2701 2.1329 0.0003 **0.2309 1.8233 0.0007** **0.0007 0.0000 0.0000**	**4.4589**			

TABLE 2(A)
THE RESULT FOR SCANNED DOCUMENT ONE WITH OUT NOISE REMOVAL



document	part	PCA

original	logo	0.8377 0.5461 0.0014
		-0.5461 0.8377 -0.0009
		-0.0017 -0.0000 1.0000

forgery	logo	**0.9999 0.0145 0.0015** **-0.0145 0.9999 -0.0000** **-0.0015 -0.0000 1.0000**

original	stamp	0.3920 0.9200 -0.0011
		0.9200 -0.3920 -0.0025
		0.0027 0.0000 1.0000

forgery	Stamp	**1.9637 0.8502 -0.0008** **0.8502 -0.5265 -0.0013** **0.0015 0.0000 1.0000**

original	Signature	0.5492 0.8357 0.0006
		0.8357 -0.5492 0.0009
		-0.0010 -0.0000 1.0000

forgery	Signature	**0.7401 -0.6725 0.0004** **0.6725 0.7401 0.0003** **-0.0005 0.0000 1.0000**

The de-noised image analysis in proposed work through using (peak signal-to-noise ratio (PSNR) and Mean Squared Error (MSE)). PSNR comparing two images one is original image and other is distorted image, small number means better de-noisy image. MSE metrics measure error between the original and the de-noised image, high value is better and refer to large difference. PSNR and MSE which are given by the following equations as [13]:

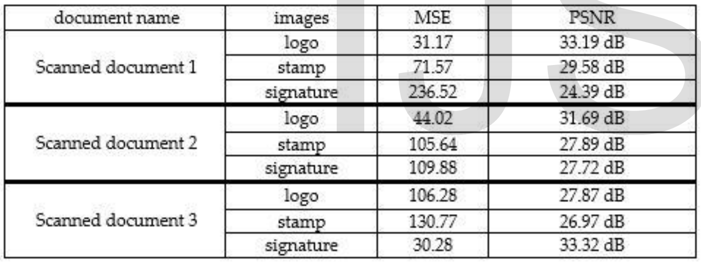
$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right) \quad (4)$$

Where R is the maximum fluctuation in the input image data type.

$$MSE = \frac{1}{MN}\sum_{I=1}^{M}\sum_{J=1}^{N}\left(X(I,J) - Y(I,J)\right)^2 \quad (5)$$

Where X (I, J) represents the original image and Y (I, J) represents the de-noised image and I and J are the pixel position of the M×N image. MSE is zero when X (I, J) = Y (I, J). Table 3 shows the PSNR values of de-noising.

TABLE 3
THE VALUES OF PSNR CALCULATIONS



document name	images	MSE	PSNR
Scanned document 1	logo	31.17	33.19 dB
	stamp	71.57	29.58 dB
	signature	236.52	24.39 dB
Scanned document 2	logo	44.02	31.69 dB
	stamp	105.64	27.89 dB
	signature	109.88	27.72 dB
Scanned document 3	logo	106.28	27.87 dB
	stamp	130.77	26.97 dB
	signature	30.28	33.32 dB

6 Conclusion

Depending on commonly using of the scanned documents, the scanned document forgery become one of well-known criminality in our life. This work focuses on methods to detect digital forgeries based digital pixel properties in the grey level, such as the max frequency intensity of pixels as the first method and the edge gradient as the second one to find some variance between the forgery and the original one, then can be detected .A crafty individual, who wants to perfect an image forgery, with time not a factor, can usually give any detection method trouble.The process of cutting the original image into several images in small sizes and specific details added an improvement of the method adopted.Another conclusion about the initial treatment paragraph and how it affected the process of (PCA) and (LDA) in reducing features. Conclusion regarding the percentage of identification of forgery of non-counterfeiting. Detection rate or identification of forgery before the process of image enhancement and noise removal and its proportion after noise removal. Time and how to reduce it by reducing the number of properties each image.

References

[1] Mahdian, B., and Saic, S.: 'Blind methods for detecting image fakery', IEEE Aerospace and Electronic Systems Magazine, 2010, 25, (4), pp. 18-24 W.-K.

[2] *Shivakumar, B., and Baboo, L.D.S.S.: 'Detecting copy-move forgery in digital images: a survey and analysis of current methods', Global Journal of Computer Science and Technology, 2010*

[3] *Zimba.M. and Xingming, S. "DWT-PCA (EVD) based copy-move image forgery detection" in International Journal of Digital Content Technology and it Applications,Vol:5,no.1,pp.251-7,2011*

[4] *Lin, S.D., and Wu, T.: 'An integrated technique for splicing and copy-move forgery image detection', in Editor (Ed.)^(Eds.): 'Book An integrated technique for splicing and copy-move forgery image detection' (IEEE, 2011, edn.), pp. 1086-1090*

[5] *Kakar, P., and Sudha, N.: 'Exposing postprocessed copy–paste forgeries through transform-invariant features', 2012*

[6] *Ahmed, A.G.H., and Shafait, F.: 'Forgery detection based on intrinsic document contents', in Editor (Ed.)^(Eds.): 'Book Forgery detection based on intrinsic document contents' (IEEE, 2014, edn.), pp. 252-256*

[7] *Abed, R.M.: 'Scanned Documents Forgery Detection Based on Source Scanner Identification', American Journal of Information Science and Computer Engineering, 2015, 1, (3), pp. 113-116*

[8] *Shahroudnejad, A., and Rahmati, M.: 'Copy-move forgery detection in digital images using affine-SIFT', in Editor (Ed.): 'Book Copy-move forgery detection in digital images using affine-SIFT' (IEEE, 2016, edn.), pp. 1-5*

[9] *Hayat, K., and Qazi, T.: 'Forgery detection in digital images via discrete wavelet and discrete cosine transforms', Computers & Electrical Engineering, 2017, 62, pp. 448-458*

[10] *Mankar, S.K., and Gurjar, A.A.: 'Image Forgery Types and Their Detection: A Review', International Journal of Advanced Research in Computer Science and Software Engineering, 2015, 5, (4), pp. 174-178*

[11] *Maintz, T.: 'Digital and medical image processing', Universiteit Utrecht, 2005*

[12] *Xiong, Y.-z., Wang, R., and Li, Z.: 'Extracting Land Use/Cover of Mountainous Area from Remote Sensing Images Using Artificial Neural Network and Decision Tree Classifications: A Case Study of Meizhou, China', in Editor (Ed.)^(Eds.): 'Book Extracting Land Use/Cover of Mountainous Area from Remote Sensing Images Using Artificial Neural Network and Decision Tree Classifications: A Case Study of Meizhou, China' (IEEE, 2010, edn.), pp. 133-136*

[13] *Ramadhan, A., Mahmood, F., and Elci, A.: 'Image denoising by median filter in wavelet domain', arXiv preprint* arXiv: 1703.06499, 2017