2018 IEEE 5th International Conference on Engineering Technologies
& Applied Sciences, 22- 23 Nov 2018, Bangkok Thailand

Improved Feature Vector of Median Filtering Residual for Image Forensics

Kang Hyeon RHEE Ilyong CHUNG
School of Electronics Engineering, Department of Computer Science
Chosun University, Gwangju 61452, Korea
E-mail: {khrhee, iyc}@chosun.ac.kr

***Abstract*—For the detection of median filtering (MF) forensics, this paper proposes an improved feature vector that consists of the refined feature vector of the MFR AR (Median Filter Residual Autoregressive Model) and an additional feature vector about the gradient and edge lines of an image. The improved nine dimensionality feature vector is trained in a SVM (Support Vector Machine) classifier for the median filtering detection (MFD) of the forged images.**

The performance of the proposed MFD scheme is measured with several types of the operated images: median filtered, unaltered and JPEG compressed, respectively.

Subsequently, the AUC (Area Under Curve) results of the proposed MFD scheme have 0.99 over on the trained SVM (Support Vector Machine) classifier. Thus, it confirmed that the grade evaluation of the proposed scheme is '*Excellent (A)*' in terms of the AUC evaluation.

Keywords: Forgery image; Median Filtering Detection (MFD); Digital Image Forensics; Median Filter Residual (MFR).

I. INTRODUCTION

In the manipulation of the forgery image, there are several alteration methods as compression, filtering, averaging, rotating, mosaic editing and up-down scaling. In particular, the median filtering (MF) is preferred among some forgers because it has the characteristics of nonlinear filtering based on order statistics [1,2].

Kang *et al*. [3] obtained autoregressive (AR) coefficients as the feature vectors for MFD via an AR model to analyze the median filter residual (the scheme called MFR AR), which is the difference between the values of the original image and those of the median filtering image. The authors analyzed an image's MFR AR; it is able to suppress image content that may interfere with the MFD.

In this paper, for the MF forensics, new MFD scheme is proposed, in which the feature vector of MFR AR is to be refined, and adding the information about gradient and edge lines of an image, by the test decision of hypothesis about the feature vectors of the median filtered image (*n*-data) and the unaltered image (*p*-data).

The rest of the paper is organized as follows: in Section 2, it briefly introduces the median filter residual (MFR) [3]. In Section 3, the newly improved feature vector is proposed for the MFD. The experimental results of the proposed MFD scheme are evaluated in Section 4. The performance evaluation compared with the prior work. Finally, the conclusion is drawn in Section 5.

II. MEDIAN FILTER RESIDUAL

In [3], the MFR (Median Filter Residual) used 10-dim. feature vector, which computed AR coefficients of the difference image $d(i,j)$ between the original $y(i,j)$ and its median filtering image $z(i,j)$. The MFR AR is formally defined as

$$\begin{aligned} d(i,j) &= \mathrm{med}_w\big(y(i,j)\big) - y(i,j) \\ &= z(i,j) - y(i,j) \end{aligned} \tag{1}$$

where (i,j) is a pixel coordinate and w is MF window size, and AR coefficients computed as

$$a_k^{(r)} = \mathrm{AR}(\mathrm{mean}\big(d^{(r)}\big)) \tag{2}$$

$$a_k^{(c)} = \mathrm{AR}(\mathrm{mean}\big(d^{(c)}\big)) \tag{3}$$

$$a_k = (a_k^{(r)} + a_k^{(c)})/2 \tag{4}$$

where r and c mean that row and column directions respectively, and k is the AR order, $1 \le k \le p$, p is the max. order. Again AR coefficients are to be the difference image by following

$$d(i,j) = -\sum_{q=1}^{p} a_k^{(r)} d(i,j-q) + \varepsilon^{(r)}(i,j) \tag{5}$$

$$d(i,j) = -\sum_{q=1}^{p} a_k^{(c)} d(i-q,j) + \varepsilon^{(c)}(i,j) \tag{6}$$

978-1-5386-7966-1/18/$31.00 ©2018 IEEE

where $\varepsilon^{(r)}(i,j)$ and $\varepsilon^{(c)}(i,j)$ are the prediction errors [4] in the row direction and column direction respectively, and q is a surrounding range of (i,j), $q < 3$.

III. New Feature Vector of the Proposed MFD Scheme

To improve the FV from the AR coefficients of the MFR [3], first examining the element of FV under each hypothesis H (Significance level: 0.05), it can observe the following

H_0: regard the same as FV values of the median filtered and unaltered image.

H_1: regard the difference as FV values of the median filtered and unaltered image.

TABLE I shows the test decision of FV of the MFR AR for the null hypothesis.

TABLE I. TEST DECISION OF THE FV OF THE MFR AR.

Image data sets	Test decision of hypothesis H
UCID	[1 1 1 1 0 0 1 0 0 0]

Consider the hypothesis in TABLE I, the FV element number of MFR AR, No. 5, 6 and 8~10 will be rejected, and No. 1~4 and 7 will be alternative. Thus, the FV length of MFR AR is reduced 5-dim.

Second, because of the FV length of MFR AR is reduced by the test decision of hypothesis H, it just adds some FV. So, it will add the FV which extracted from the information about gradient and edge lines of an image.

In this paper, it proposes the added FV (called *aFV*) from the following

Added FV:

aFV1: Gradient magnitude of an image by Prewitt gradient operator.

aFV2: Gradient direction of an image by Prewitt gradient operator.

aFV3: Average value of a binary image from except 16-pixel objects by Canny edge operator.

aFV4: Average value of a binary image from except 32-pixel objects by Canny edge operator.

TABLE II. TEST DECISION OF THE *aFV* OF THE PROPOSED SCHEME.

Image data sets	Test decision of hypothesis H
UCID	[1 1 1 1]

Third, the *aFV* No. 1~4 should process the test decision of hypothesis H like the MFR AR. Table II shows the test decision of the *aFV* of the proposed scheme for the null hypothesis.

Consider the hypothesis in TABLE II, the *aFV* element number of the proposed scheme, No. 1~4 will be an alternative. Thus, the *aFV* length of the proposed scheme is defined 4-dim.

Lastly, the reduced FV of the MFR AR and the *aFV* of the proposed scheme is combined with 9-dim. by the test decision of hypothesis H between the median filtered and unaltered images.

IV. Experimental Result

In this section, first, it describes the experimental methodology. Second, the experimental results of the proposed MFD scheme compare to [3] to verify the performance.

A. Experimental Methodology

The new 9-dim. feature vector inputted to a SVM classifier for the training of the classification of the MF image and the other operated image types. C-SVM with Gaussian kernel is employed as the classifier

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \qquad (\gamma > 0).$$

Moreover, it trained in a SVM classifier with five-fold cross-validation in conjunction with a grid search for the best parameters of C and γ in the multiplication grid [5].

$$(C, \gamma) \in \{(2^i, 2^j) | 4 \times i, 4 \times j \in Z\}.$$

The searching step size of (i, j) is 0.25 then those parameters are used to get the classifier model on the training set similar to [1-3], and the training images prepared following image database:

- The UCID image database that consists of 1,338 images [6].

The images of the UCID database are 8-bit grayscale and used in experiments.

B. Experimental Result

Fig. 1 presents a Cut-Paste forgery example image. The forgery image experiments on the trained SVM classifier, and the ROC curves and AUCs (Area Under ROC Curve) [7] are shown in Fig. 2 (the MFR AR [3]) and Fig. 3 (The proposed MFD scheme), respectively, where (MF3: median filtering (w=3 x 3) and JPG90: JPEG compression (QF=90)). Also, the forgery detection is examined each image block size 32 x 32 and 64 x 64, respectively.

The AUC is compared between the MFR AR [3] and the proposed scheme in TABLE III.

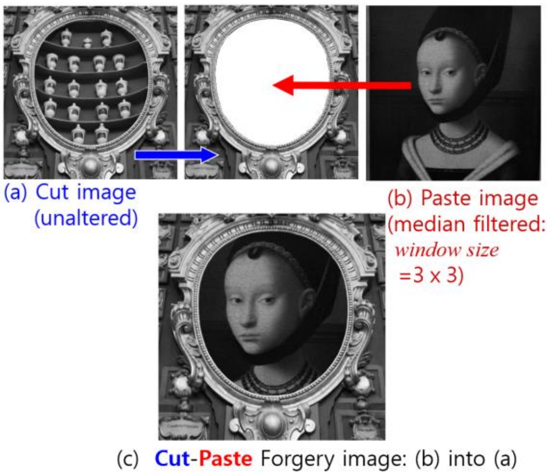

Fig. 1. Cut-Paste forgery image example.

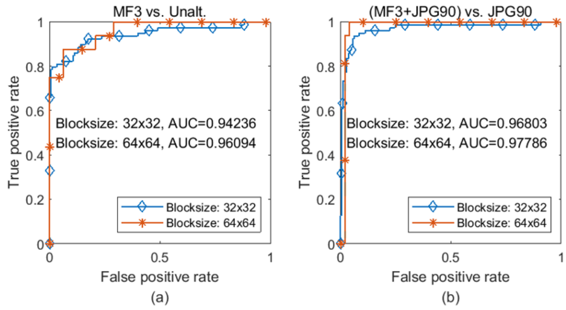

Fig. 2. ROC curves of the MFR AR [3].

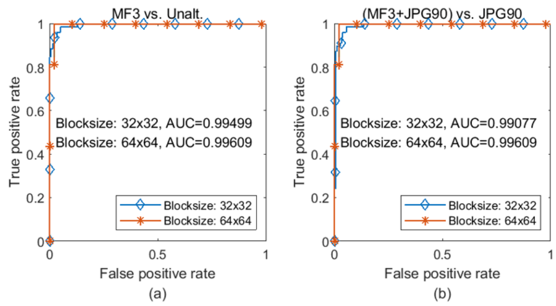

Fig. 3. ROC curves of the proposed MFD scheme.

TABLE III. COMPARISON OF $\mathcal{AUC}$ BETWEEN THE MFR AR ($\mathcal{A}$) AND THE PROPOSED SCHEME ($\mathcal{B}$).

BS: Test image block size $\mathcal{BS}32$: 32 x 32, $\mathcal{BS}64$: 64 x 64

1: MF3 vs. Unaltered image

2: (MF3 + JPG90) vs. JPG90 image

$\mathcal{AUC}$	*1*		*2*	
BS	$\mathcal{BS}32$	$\mathcal{BS}64$	$\mathcal{BS}32$	$\mathcal{BS}64$
$\mathcal{A}$	0.942	0.962	0.968	0.977
$\mathcal{B}$	**0.995**	**0.996**	**0.991**	**0.996**

The resulted AUCs by the sensitivity (TP: True Positive rate) and 1-specificity (FP: False Positive rate) of the proposed MFD scheme is 0.99 over.

However, the proposed MFD 9-dim. feature vector is improved than the MFR AR [3], and it confirms that our scheme rated as '*Excellent* (A)' in terms of the AUC evaluation.

In Fig. 4 and Fig.5, the detected paste area marked the red color pattern.

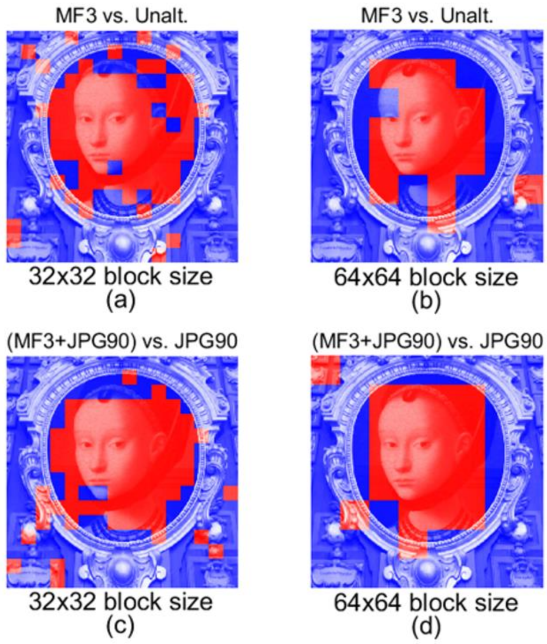

Fig. 4. Paste area detection (red color) in the forgery image by the MFR AR [3].

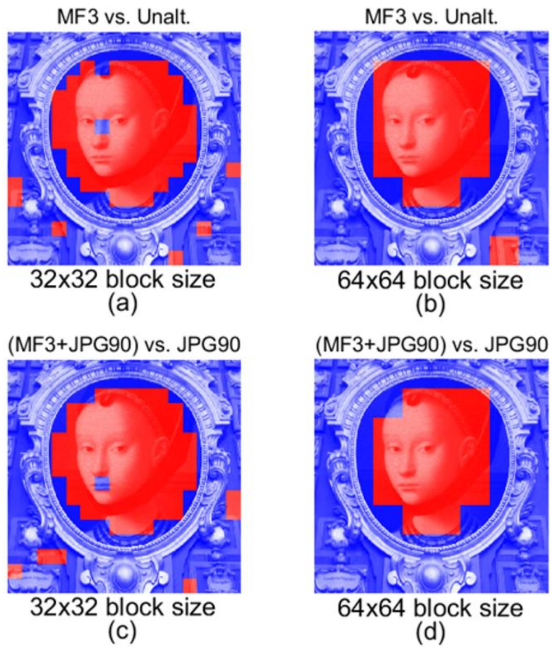

Fig. 5. Paste area detection (red color) in the forgery image of the proposed MFD scheme .

V. CONCLUSION

In this paper, the median filtering forensic detection scheme of the forgery image is proposed. The performance of the improved feature vector: *the refined MFR feature vector and the added feature vector* in the proposed MFD scheme is excellent. In the proposed scheme, the improved feature vector could be applied to the MF image forensics. So this will serve as the further research area of the variety of forgery images.

ACKNOWLEDGMENT

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW) supervised by the IITP (Institute for Information & communications Technology Promotion) (2017-0-00137)

REFERENCES

[1] Kang Hyeon RHEE, "Median filtering detection using variation of neighboring line pairs for image forensics," *SPIE, Journal of Electronic Imaging*, 25(5), pp. 53039-1~053039-13, Sep./Oct., 2016.

[2] Kang Hyeon RHEE, "Median Filtering detection based on variations and residuals in image forensics," *Turkish Journal of Electrical Engineering & Computer Sciences*, (2017) 25: 3811-3826, 2017.

[3] Xiangui Kang, Matthew C. Stamm, Anjie Peng, and K. J. Ray Liu, "Robust Median Filtering Forensics Using an Autoregressive Model," *IEEE Trans. on Information Forensics and Security*, vol. 8, no. 9, pp. 1456-1468, Sept. 2013.

[4] S. M. Kay, Modern Spectral Estimation: Theory and Application. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[5] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Syst. and Technol.*, pp. 2:27:1–27:27, 2011.

[6]http://homepages.lboro.ac.uk/~cogs/datasets/ucid/ucid.html (2016.1.15)

[7] http://gim.unmc.edu/dxtests/roc3.htm (March 2016)