Blind Integrity Verification of Medical Images

H. Huang, G. Coatrieux, *Member, IEEE*, H.Z. Shu, *Senior Member, IEEE*, L.M. Luo, *Senior Member, IEEE*, Ch. Roux, *Fellow, IEEE*

***Abstract*—This work presents the first method of digital blind forensics within the medical imaging field with the objective to detect whether an image has been modified by some processing (e.g. filtering, lossy compression and so on). It compares two image features: the Histogram statistics of Reorganized Block-based Discrete cosine transform coefficients (HRBD), originally proposed for steganalysis purposes, and the Histogram statistics of Reorganized Block-based Tchebichef moments (HRBT). Both features serve as input of a set of SVM classifiers built in order to discriminate tampered images from original ones as well as to identify the nature of the global modification one image may have undergone. Performance evaluation, conducted in application to different medical image modalities, shows that these image features can help, independently or jointly, to blindly distinguish image processing or modifications with a detection rate greater than 70%. They also underline the complementarity of these features.**

***Index Terms*— classification, image moments, blind forensics**

I. Introduction

WITH the development of multimedia and communication technologies in complex, distributed and cooperative domains like telemedicine, trust one can have in medical data becomes critical. Medical images, for which advanced editing and sharing tools are proposed, are also concerned. Global image processing like filtering can be used by the physician during his or her interpretation to raise up some specific pieces of image information [1]. But at the same time some specific information is put in evidence, one other is masked or definitively lost if the process is not reversible. Lossy image compression (e.g. JPEG…) acts similarly so as to make possible image sharing on low bit rate channels like those encountered in telemedicine or to gain in storage capabilities. Notice that studies are regularly conducted in order to determine the good compromise between compression rate and preservation of the image quality for the diagnosis [2][3]. Depending on their extent or strength, these processes may endanger the diagnosis value of images, by inducing loss of critical information. Even though the medical image standard DICOM (Digital Imaging and Communications in Medicine - medical.nema.org) traces such modifications by means of "indicators" in the image file header, it must be known that not all medical image software or devices are fully DICOM compliant introducing interoperability and data integration issues [4-6]. For instance, only some commercial implementations declare their compliance with the security profiles of DICOM [7]. These profiles establish for example how to verify the integrity of an image with digital signatures (see DICOM Part 15 [8]). As a consequence, images can be shared without proofs of their authenticity. Thus, how can we be sure that an image received by an hospital is really what it claims to be? Is it a pristine sensor image or a post-processed one? Beyond interoperability issues, stand malevolent acts. It is not impossible modifying an image and its header using a third-party software in order to mask an error. If a digital signature will allow detecting such a situation, it however gives no clues about the nature of the modification the image undergone and consequently no idea about which information has been lost. This is of major concern in a medico legal framework when the image comes as evidence. As illustrated, one can no longer take the authenticity of images for granted and there is a need to inform practitioners that an image is not what it claims to be (e.g. an original image or a post-processed image). These are the objectives we pursue in this work.

Among the different solutions for validating the authenticity of an image, "blind forensics" techniques have recently attracted attention. They look at the detection of an image modification without any a priori knowledge about the image under observation (e.g. a signature stored or shared with the image). A large group of “blind forensics” methods are based on classifier mechanisms built on some image features so as to recognize modification footprints. Most image feature sets from the literature have been first proposed for the steganalysis purpose [9-11]; that is detecting the presence of secret message embedded in images; and, next for image authentication [12-15] but only in the case of natural images, i.e. images for the general public. In fact, steganography acts like image processing. It affects more or less the image content while not carrying out perceptible distortions [15]. Recently, for the steganalysis purpose only, Liu *et al*. [16] designed an image feature set from the Histogram statistical properties of Reorganized Block-based Discrete cosine transform coefficients (HRBD) which outperforms other strategies [16]. In this work, we experiment and extend their approach within the framework of medical image authentication while working with the Tchebichef moments in addition to the DCT coefficients. We thus propose a new image feature set, the HRBT. Moreover, in this work, we decided to go further than simply detecting an image undergone or not some “global” image modification. We also aim at discriminating the nature of the modification. The rest of this paper is organized as follows.

H. Huang, G. Coatrieux and Ch. Roux are with the Institut Mines-Telecom; Telecom Bretagne; Unite INSERM 1101 Latim, Technopole Brest-Iroise, CS 83818, 29238 Brest Cedex 3 France (e-mail: hh.sophia@gmail.com, {gouenou.coatrieux, christian.roux}@telecom-bretagne.eu).

H.Z. Shu and L.M. Luo are with the LIST Southeast University, China (e-mail: {shu.list, luo.list}@seu.edu.cn).

Section 2 remembers the basic principles of classifier based blind forensic mechanisms and presents the HRBD and HRBT image feature sets. Section 3 provides some experimental results obtained independently with HRBD and HRBT features and with their combination on different medical image modalities: X-Ray, Ultrasound and Magnetic Resonance Imaging. Conclusions are given in section 4.

II. Blind Image Forensics

A. Basic Principles

To catch up evidence left by image processing, one approach consists in training a classifier which uses as input some image features normally altered by image modifications. Once the classifier trained, one just has to extract these features from one image under investigation and provide them to the classifier for analysis. Efficiency of such an approach largely depends on: i) the design of proper image features and, ii) the way the classifier is built. In this work, in order to only evaluate image feature performance, we use Support Vector Machines (SVM) [17]. This choice stands on the fact that SVM have shown superior classification performances in many applications [18][19].

In our context, our primary objective is to distinguish modified images from original ones (i.e. not modified). To achieve this goal, we have trained different binary SVM classifiers or "detectors" (e.g. original images *vs.* JPEG modified images, original images *vs.* filtered modified images, and so on) using the HRBT or HRBD features. An image is declared unauthentic, if at least one of these classifiers notifies it. Secondly, for the purpose of determining the type of the modification, we build a multi-class classifier based on One-Versus-One (OVO) binary classifiers, each of which discriminates images modified accordingly to two kinds of possible modifications (e.g. JPEG *vs.* Filtering, JPEG *vs.* Contrast adjustment, Contrast adjustment *vs.* Scaling and so on). By analyzing the responses of these classifiers, a multi-class conclusion is drawn. Among the different strategies for combining decisions of binary classifiers, the Max-Wins Voting (MWV) is one of the most commonly used approaches [20]. MWV assigns an instance to a class which has the largest votes from all binary classifiers.

B. HRBD and HRBT Image Feature sets

HRBD [8] and HRBT features are extracted following the same strategy. Their difference stands in the image "transformed" coefficients considered: the DCT coefficients for HRBD and, the Tchebichef moments for HRBT. These features aim at carrying out the variations of the histograms of groups of such DCT coefficient or moments. Notice that some of these features have been previously experimented in the spatial [21] and wavelet [22][23] domains.

The first group of features we use, has been proposed by Shi *et al.* [23]. They correspond to the statistical moments of the Discrete Fourier Transform of one Histogram (DFTH):

$$M_n = \left(\sum_{k=0}^{K/2} k^n |H(k)|\right) \Big/ \left(\sum_{k=0}^{K/2} |H(k)|\right), \quad n = 1,2,3 \quad (1)$$

where $H(k), |H(k)|$ denote the DFTH at the frequency k and its magnitude respectively, and K corresponds to the frequency dynamic. The second group of features we also consider, has been suggested by Wang *et al.* [24]. They improve steganalysis performances by working on filtered version of the DFTH. They suggest three distinct features $f_i, i = 1 \dots 3$

$$f_1 = \sum_{k=0}^{K/2} |H(k)| \cdot \sin(\pi k/K) \quad (2)$$

$$f_2 = \sum_{k=0}^{K/2} |H(k)| \cdot \sin^2(\pi k/K) \quad (3)$$

$$f_3 = \sum_{k=0}^{K/4} |H(k)| \cdot \sin(\pi k/K) \quad (4)$$

In order to be more image content independent, Shi *et al.* [15] proposed to extract these features from the image and also from its prediction error image, which is the difference between the image and its predicted version. The prediction algorithm, we exploit in the sequel, is expressed as [25]:

$$e = x - \hat{x} = \begin{cases} max(a,b) & c \leq min(a,b) \\ min(a,b) & c \geq max(a,b) \\ a+b-c & otherwise \end{cases} \quad (5)$$

where *e* is the prediction error of the pixel *x*; $\hat{x}$ is the predicted value of x; x, a, b and c are the 4 pixels in a 2×2 pixel block.

From that standpoint, Liu *et al.* [16] suggest to extract these DFTH statistics from re-organized block based DCT coefficients; providing a set of features we named HRBD.

Herein, in the framework of image authentication, we propose to substitute DCT coefficients by the Tchebichef moments of the image. These moments are the simplest discrete orthogonal moments, widely used in digital image processing such as image reconstruction and pattern recognition. The $(n+m)^{th}$ order Discrete Tchebichef moment T_{nm} is defined by [26]:

$$T_{nm} = 1/\rho(n,N)\rho(m,N) \cdot \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t_n(x) t_m(y) f(x,y) \quad (6)$$

where $t_n(x)$ and $t_m(y)$ are the Tchebichef orthogonal polynomials and $\rho(n,N)$ weighted values:

$$t_n(x) = n! \sum_{k=0}^{n} (-1)^{n-k} \binom{N-1-k}{n-k} \binom{n+k}{n} \binom{x}{k} \quad (7)$$

$$\rho(n,N) = (2n)! \binom{N+n}{2n+1} \quad (8)$$

Notice that, it is possible to derive the DCT basis from the discrete Tchebichef moments [27] and that some experimental results given in [28] illustrate close performances between DCT coefficients and Tchebichef moments for lossy image compression and image reconstruction. However, in some cases, Tchebichef moments have shown better behavior especially for images with sharp boundaries and regular texture [20]. This motivates and makes us expect achieving better performances with the two sets of features.

The procedure for extracting HRBT features is the same than for HRBD features presented in [16]. In a first time, the image is divided into non-overlapping blocks of u×u pixels - in the sequel u=8 - and for each, Tchebichef moments up to the $((u-1)+(u-1))^{th}$ order are computed leading to u×u moments values: $T_{pq}, p = 0, \dots, u-1; q = 0, \dots, u-1$. Then, each block of u×u moments is partitioned into an L-scale wavelet-like tree of $3L + 1$ subbands, where $u = 2^L$. At last, moments belonging to the same sub-band in each block are clustered so as to form an L-scale coefficients tree for the whole image. DFTH statistics are then extracted from each of these "subbands" $\{G_0, G_1, \dots, G_{3L}\}$.

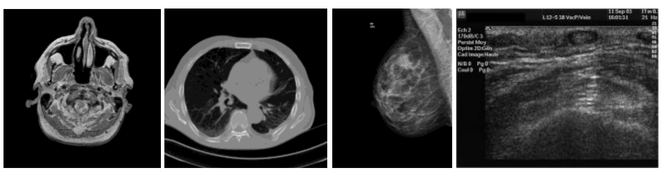
III. Experimental Results

A. *Image test sets and modifications*

Four test sets of images issued from different medical image modalities (see Fig. 1) were considered:

- Magnetic resonance images (MRI) of the head: 120 images of 12 bit depth and of 256 × 256 pixels from 3 patients;
- 200 abdomen CT images from 4 patients with 512 × 512 pixels coded on 12 bits;
- X-Ray Imaging: 162 mammograms of 4740×3540 pixels coded on 12 bits from multiple patients;
- Ultrasound imaging (Echo): 52 images of 576 × 690 pixels and 8-bit depth of thrombosis from 3 patients.

The modifications we have considered in these experiments are: contrast and brightness adjustment, Gaussian filtering, scaling, Laplacian filtering, JPEG and JPEG2000 lossy compression and histogram equalization. Table I gives the parameters used for each of these modifications.



(a) (b) (c) (d)

Fig. 1. Image samples from our test sets: (a) MRI (b) CT image (c) X-Ray image d) Echography.

TABLE I Image manipulation and their parameters

Modification	Values of parameters				
Scaling up(γ_s %)	1	5	10	25	50
Laplacian filter(α)	0.1	0.3	0.5	0.7	0.9
Deviation of Gaussian filter (σ)	0.3	0.5	1.0	2.0	3.0
Contrast enhancement rate (γ_c %)	1	5	8	10	
Brighten rate (γ_b %)	2	5	8	10	
Quality factor(Q)	95	85	80	75	60
Compression rate JP2K (γ_j)	2:1	5:1	10:1	20:1	50:1
Histogram equalization					

For each modality, the different image test sets were divided into two groups for training and testing our classifiers (see section II-A). By next, we give the average results achieved with classifiers that have been trained several times (at least 10 times) with different fold-cross validation (i.e. training and testing sets are randomly selected at each trial). SVM Classifiers were used so as to only evaluate performance of HRBD and HRBT features or of their combination. Notice that features are extracted from one single 128×128 pixel block centered in each image, while considering 13 "subbands" or group of coefficients. Thus, in the following, an image is summed up by an HRBT feature vector (resp. an HRBD feature vector) of 156 components or by a vector of 312 components if HRBT and HRBD features are used jointly. Notice also that the parameters of the SVM were selected from their Receiver Operating Characteristics (ROC) curves.

B. *Performances Evaluation*

As mentioned above, one first objective is to discriminate original images from modified ones. Once an image is declared unauthentic, a multi-class classifier is exploited to distinguish the nature of the modification. Due to paper length limitation, only the detection rate is used as performance indicator; rates that are listed in Table II and Table III for HRBD and HRBT features and for their combination.

Table II Detection rates - modified *vs.* non-modified images

Detection rate (%)	MRI	Mammography	CT	Echography
HRBT	78.51	84.84	83.21	81.65
HRBD	78.26	84.49	82.47	84.67
HRBT&HRBD	79.13	85.67	85.06	85. 87

From Table II, it can be viewed that both HRBD and HRBT features can be used for image modification detection. They have similar performances with detection rates about 80% whatever the image modality. From our experiments, HRBT features perform better than HRBD features for X-ray images (mammography and CT images). According to our previous comments (see section II), one reason may stands in the fact that X-Ray images have sharp boundaries. Notice also that the false positive and negative detection rates achieved by our binary classifiers are similar and lower than 20%, which is a good tradeoff. The combination of HRBD and HRBT into a single feature set gives better performance. This is in agreement with a result shown by Bayram *et al.* [15]: detection performance with several sets of image features are better or at least equal to those obtained considering these feature sets independently. However, the gain is rather small and can be explained by the close mathematical relationship of HRBT and HRBD features.

In Table III, we give the correct identification rates of the modification considering images declared unauthentic at the previous stage. In general, it can be seen that the modification can be identified with a rate higher than 70% for HRBT and HRBD or than 77% if these features are considered jointly, except for contrast and brightness adjustment. One reason may stand in the fact that these two modifications confuse our multiclass-classifier. If one of these modifications is omitted in the experiment, the detection rate is higher (greater than 78%). Table III also points out that HRBD features perform better than the HRBT for MRI and Mammography images but not for CT and echography images. This underline that HRBT and HRBD features put in evidence different pieces of image information that are more or less sensitive to image modifications. It appears that HRBD and HRBT are complementary features. As example, HRBD perform better for identifying JPEG images but not for CT images where HRBT features provide better results. This is also supported by the better detection rates we obtain when HRBD and HRBT are exploited simultaneously.

From the above experiments, where image features are extracted from a single block of 128x128 pixels centered into the image, one can expect detecting image processing applied locally. However, detection rates will decrease with the size of the block on which the analysis is conducted. For instance, in the case classifiers are trained based on features extracted from one 32x32 pixel block (instead of 128x128), the detection rates for discriminating original images (or original pixel blocks) from modified ones, fall in the range 74%-81% whatever the image modality. These rates are about 5% less than with a block of 128x128 pixels (see table II).

Actually, our system keeps limited to the detection or identification of “a priori known” modifications. In the case of an “unforeseen attack”, this one will be identified as the modification that leaves the most similar footprint into the image. Nevertheless, our solution can be easily updated. It does not require to recompute and store a new image signature.

Table III MULTI-CLASS CLASSIFIER DETECTION RATES BASED ON HRBT/HRBD/HRBT&HRBD FEATURES

Detection rate (%)	MRI HRBT/ HRBD/ HRBT& HRBD	Mammo. HRBT/ HRBD/ HRBT& HRBD	CT HRBT/ HRBD/ HRBT& HRBD	Echography HRBT/ HRBD/ HRBT& HRBD
JPEG2K	78.53/83.56/ 84.99	71.38/82.59/ 85.56	86.38/90.94/ 92.93	93.75/85.05/ 96.92
JPEG	79.47/99.01/ 98.37	83.84/89.75/ 98.02	89.17/63.37/ 93.60	98.56/98.08/ 99.78
Gaussian filtering	75.17/77.00/ 77.62	81.60/86.79/ 89.14	78.04/69.40/ 79.30	86.15/76.72/ 87.85
Laplacian filtering	100/100/ 100	100/100/ 100	100/100/ 100	100/100/ 100
Scaling	72.31/79.23/ 80.69	81.00/86.17/ 89.75	81.39/75.52/ 83.60	86.73/86.57/ 91.92
Brighten	68.56/73.54/ 74.14	65.87/70.27/ 73.12	73.52/57.15/ 76.50	75.73/72.93/ 77.52
Contrast	59.85/82.08/ 83.23	59.56/62.52/ 66.64	73.14/78.15/ 79.38	63.78/67.71/ 69.23
Histeq.	100/100/ 100	100/100/ 100	100/100/ 100	100/100/ 100

IV. CONCLUSION

In this paper, we have shown that blind forensics approaches initially proposed for general public images, can also be used in medical imaging. Without taking care of the signal specificities, good detection rates are already achieved. We also shown that HRBD features, originally proposed for natural image steganalysis, complemented with new HRBT features can serve blind detection of global image modification and, furthermore, for determining the nature of the modification one image may have undergone. Depending on the image modality and as well as on the type of the modification, HRBT and HRBD features are complementary. At least, this experiment points out that image moment theory can be exploited for verifying blindly the integrity of medical images. Future works will focus on identifying more appropriate image moments as well as on the detection of local image modification.

REFERENCES

[1] A.P. Dhawan, Medical Image Analysis, Second Edition, Institute of Electrical and Electronics Engineers, Inc., John Wiley Publications and IEEE Press, 2011.

[2] N. Gupta, M. N. S. Swamy, and E. Plotkin, “Despeckling of medical ultrasound images using data and rate adaptive lossy compression,” Medical Imaging, IEEE Transactions on, vol. 24, no. 6, pp. 743–754, 2005.

[3] K. Chen and T. V. Ramabadran, “Near-lossless compression of medical images through entropy-coded DPCM,” Medical Imaging, IEEE Transactions on, vol. 13, no. 3, pp. 538–548, 1994.

[4] A. G. Farman, “Raising standards: digital interoperability and DICOM,” Oral Surg Oral Med Oral Pathol Oral Radiol Endod, vol. 99, no. 5, pp. 525–526, May 2005.

[5] H. J. Oosterwijk, “DICOM: your guarantee to interoperability?,” Proceedings of SPIE, vol. 3035, no. 1, pp. 165–169, May 1997.

[6] J. Broda, “Interoperability and the healthcare enterprise,” online EHM Magazine, GDS Publishing, issue 1, http://www.executivehm.com/article/Interoperability-and-the-healthcare-enterprise/,

[7] L.O. Kobayashi, S.S. Furuie, “Proposal for DICOM multiframe medical image integrity and authenticity,” J Digit Imaging. 2009 Mar;22(1):71-83.

[8] “Digital Imaging and Communications in Medicine (DICOM) Part 15: Security and System Management Profiles.” National Electrical Manufacturers Association, 2004.

[9] I. Avcibas, N. D. Memon, and B. Sankur, “Steganalysis of watermarking techniques using image quality metrics,” in *Security and Watermarking of Multimedia Contents III*, San Jose, CA, USA, 2001, vol. 4314, pp. 523-531.

[10] I. Avcibaş, M. Kharrazi, N. Memon, and B. Sankur, “Image steganalysis with binary similarity measures,” *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 2749-2757, 2005.

[11] H. Farid, “Detecting hidden messages using higher-order statistical models,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, vol. 2, p. II-905-II-908 vol.2.

[12] X. Guo and T.-ge Zhuang, “Lossless Watermarking for Verifying the Integrity of Medical Images with Tamper Localization,” *Journal of Digital Imaging*, vol. 22, no. 6, pp. 620-628, Dec. 2009.

[13] I. Avcibas, S. Bayram, N. Memon, M. Ramkumar, and B. Sankur, “A classifier design for detecting image manipulations,” presented at the Image Processing, 2004. ICIP ’04. 2004 International Conference on, 2004, vol. 4, pp. 2645-2648 Vol. 4.

[14] H. Farid and S. Lyu, “Higher-order Wavelet Statistics and their Application to Digital Forensics,” presented at the Computer Vision and Pattern Recognition Workshop, 2003. CVPRW ’03. Conference on, 2003, vol. 8, p. 94.

[15] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, “Image manipulation detection,” *Journal of Electronic Imaging*, vol. 15, no. 4, pp. 041102-17, Oct. 2006.

[16] S. Liu, L. Ma, H. Yao, and D. Zhao, “Universal Steganalysis Based on Statistical Models Using Reorganization of Block-based DCT Coefficients,” in *2009 Fifth International Conference on Information Assurance and Security*, 2009, pp. 778–781.

[17] L. Wang, *Support vector machines: theory and applications*. Springer, 2005.

[18] D. Fu, Y. Q. Shi, and W. Su, “Detection of Image Splicing Based on Hilbert-Huang Transform and Moments of Characteristic Functions with Wavelet Decomposition,” *Digital Watermarking*, vol. 4283, pp. 177-187, 2006.

[19] D. J. Sebald and J. A. Bucklew, “Support vector machine techniques for nonlinear equalization,” *Signal Processing, IEEE Transactions on*, vol. 48, no. 11, pp. 3217-3226, 2000.

[20] U. H. . Kressel, “Pairwise classification and support vector machines,” in *Advances in kernel methods*, 1999, pp. 255–268.

[21] J. J. Harmsen and W. Pearlman, “Steganalysis of additive noise modelable information hiding,” in *Proc. SPIE Electronic Imaging*, 2003, vol. 5022.

[22] G. Xuan et al., “Steganalysis Based on Multiple Features Formed by Statistical Moments of Wavelet Characteristic Functions,” in *Information Hiding*, 2005, pp. 262-277.

[23] Y. Q. Shi et al., “Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network,” in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, 2005, p. 4.

[24] Y. Wang and P. Moulin, “Statistical modeling and steganalysis of DFT-based image steganography,” in *Proceedings of SPIE*, 2006, vol. 6072, pp. 14–24.

[25] M. J. Weinberger, G. Seroussi, and G. Sapiro, “LOCO-I: A low complexity, context-based, lossless image compression algorithm,” in *Proc. 1996 Data Compression Conference*, 1996, pp. 140–149.

[26] R. Mukundan, S. H. Ong, and P. A. Lee, “Image analysis by Tchebichef moments,” *Image Processing, IEEE Transactions on*, vol. 10, no. 9, pp. 1357-1364, 2001.

[27] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete Cosine Transfom,” *Computers, IEEE Transactions on*, vol. 23, no. 1, pp. 90-93, 1974.

[28] R. Mukundan, “Transform coding using discrete Tchebichef polynomials,” in *International conference Visualization, Imaging and Image processing*, 2006.