Volume 8, No. 5, May-June 2017
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

A New Copy-Move Image Forgery Detection Based on DCT

Madhu
Research Scholar (Computer Science& Eng.)
BBAU University, Lucknow,Uttar Pradesh, India

Jitendra Kurmi
Assistant Professor (Computer Science& Eng.)
BBAU University,Lucknow,Uttar Pradesh,India

Abstract: Today manipulation of digital images has become easy due to powerful computers, advanced photo-editing software packages and high resolution capturing devices. Verifying the integrity of images and detecting traces of tampering without requiring extra prior knowledge of the image content. When image software tools develop, images are modified by different purposes. The authenticity of image or image forensics therefore is important and has challenged the researchers and scientists in fields related to image processing. Many techniques are proposed for image security using watermarking or digital signatures in which the information of original images and embedded security codes are known in advance. However, not all images are provided the original information and so an arbitrary image cannot be confirmed a forged image or not by naked eyes. A blind/passive problem is considered and Copy-Move detection is one kind of this problem. CopyMove is a counterfeit manipulation in which a region is cut and pasted into another place in the same image to hide any information or make a fake object. Existing techniques for copy-move forgery detection can be classified into two categories, i.e. block-based and Feature based methods. The block-based methods usually extract features from overlapping blocks of the image, a number of features have been proposed for forgery detection. With a large number of image blocks, the time consumption of block-based methods is high, especially with the increasing size of images. In addition, most of these methods show lack of robustness against generic geometric transformations, i.e. affine and projective transforms in particular. This work we Study and analyze various image forgery techniques and Propose a new DCT based method for forgery identification.

Keywords: Passive Forensics, Copy-Move Forgery Detection, DCT

1.INTRODUCTION

Image Forgery detection

We are obviously residing into an majority where we are uncovered according to a inconceivable order over visible imagery. While we may hold traditionally had self assurance into the morality about it imagery, today's digital technology has begun according to erode this trust. From the tabloid magazines in conformity with the fashion enterprise or among mainstream media outlets, scientific journals, political campaigns, courtrooms, yet the picture hoaxes that bank of our email in-boxes, doctored images are acting including a flourishing frequency yet sophistication. Over the past five years, the subject about digital forensics has arisen in conformity with assist repair partial believe after digital images. Here I review the administration of the artwork among it instant yet exciting field. Digital watermarking has been proposed as a capacity by which an picture perform stay authenticated (see, because example, and because of ordinary surveys). The downside on this strategy is that a watermark have to be inserted at the epoch of recording, who would control this strategy in accordance with mainly outfitted digital cameras. In distinction in imitation of it approaches, dead strategies because of photo forensics operate among the penurity concerning any watermark and signature. These strategies work on the allowance as even though digital forgeries may additionally depart no visual clues to that amount indicate tampering, she may additionally deflect the underlying data over an image. The set over photo forensic equipment can lie round grouped of five categories:

1) pixel-based techniques that realize statistical anomalies brought at the pixel level;
2) format-based strategies that leverage the statistical correlations brought by using a precise lossy depth scheme;
3) camera-based strategies so much take advantage of artefacts added by using the digital camera lens, sensor, yet on-chip post-processing;
4) physically primarily based techniques so much apparently mannequin yet detect anomalies between the three-d interplay within bodily objects, light, or the camera; then
5) geometric-based methods that fulfil measurements of objects within the world and theirs positions kin according to the camera.

2.LITERATURE SURVEY

Lee, Barry B., et al. (1990) [1]In this paper, we measured the sensitivity of macaque ganglion cells to luminance and chromatic sinusoidal modulation. Phasic ganglion cells of the magnocellular pathway (M-pathway) were the more sensitive to luminance modulation, and tonic ganglion cells of the parvocellular pathway (P-pathway) were more sensitive to chromatic modulation. With decreasing retinal illuminance, phasic ganglion cells' temporal sensitivity to luminance modulation changed in a manner that paralleled psychophysical data. The same was true for tonic cells and chromatic modulation. Taken together, the data suggest strongly that the cells of the M-pathway form the physiological substrate for detection of luminance modulation and the cells of the P-pathway the substrate for detection of chromatic modulation. However, at high light levels, intrusion of a so-called luminance mechanism near 10 Hz in psychophysical detection of chromatic modulation is probably due to responses in the M-pathway, arising primarily from a nonlinearity of cone summation. Both phasic and tonic ganglion cells responded to frequencies higher than can be psychophysically detected. This suggests that central mechanisms, acting as low-pass filters, modify these cells' signals, though the corner frequency is lower for

© 2015-19, IJARCS All Rights Reserved

the P-pathway than for the M-pathway. For both cell types, the response phase at different frequencies was consistent with the cells' description as linear filters with a fixed time delay.

Cortes, Corinna et al. (1995) [2]In this paper, the support-vector network is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high-dimension feature space. In this feature space a linear decision surface is constructed. Special properties of the decision surface ensures high generalization ability of the learning machine. The idea behind the support-vector network was previously implemented for the restricted case where the training data can be separated without errors. We here extend this result to non-separable training data. High generalization ability of support-vector networks utilizing polynomial input transformations is demonstrated. We also compare the performance of the support-vector network to various classical learning algorithms that all took part in a benchmark study of Optical Character Recognition.

Zhang, Guangcheng et al. (2004) [3] In this paper presents a novel approach for face recognition by boosting statistical local features based classifiers. The face image is scanned with a scalable sub-window from which the Local Binary Pattern (LBP) histograms [14] are obtained to describe the local features of a face image. The multi-class problem of face recognition is transformed into a two-class one by classifying every two face images as intra-personal or extra-personal ones [9]. The Chi square distance between corresponding Local Binary Pattern histograms of two face images is used as discriminative feature for intra/extra-personal classification. We use AdaBoost algorithm to learn a similarity of every face image pairs. The proposed method was tested on the FERET FA/FB image sets and yielded an exciting recognition rate of 97.9%.

Ng, Tian-Tsong et al. (2004) [4]In this paper, image splicing is a simple process that crops and pastes regions from the same or separate sources. It is a fundamental step used in digital photomontage, which refers to a paste-up produced by sticking together images using digital tools such as Photoshop. Examples of photomontages can be seen in several infamous news reporting cases involving the use of faked images. Searching for technical solutions for image authentication, researchers have recently started development of new techniques aiming at blind passive detection of image splicing. However, like most other research communities dealing with data processing, we need an open data set with diverse content and realistic splicing conditions in order to expedite the progresses and facilitate collaborative studies. In this report, we describe with details a data set of 1845 image blocks with a fixed size of 128 pixels x 128 pixels. The image blocks are extracted from images in the CalPhotos collection [CalPhotos'00], with a small number of additional images captured by digital cameras. The data set include about the same number of authentic and spliced image blocks, which are further divided into different subcategories (smooth vs. textured, arbitrary object boundary vs. straight boundary).

Johnson, Micah K., et al. (2006) [5]In this paper, virtually all optical imaging systems introduce a variety of aberrations into an image. Chromatic aberration, for example, results from the failure of an optical system to perfectly focus light of different wavelengths. Lateral chromatic aberration manifests itself, to a first-order approximation, as an expansion/contraction of color channels with respect to one another. When tampering with an image, this aberration is often disturbed and fails to be consistent across the image. We describe a computational technique for automatically estimating lateral chromatic aberration and show its efficacy in detecting digital tampering.

Sokolova, Marina et al. (2006) [6]In this paper, different evaluation measures assess different characteristics of machine learning algorithms. The empirical evaluation of algorithms and classifiers is a matter of on-going debate among researchers. Most measures in use today focus on a classifier's ability to identify classes correctly. We note other useful properties, such as failure avoidance or class discrimination, and we suggest measures to evaluate such properties. These measures – Youden's index, likelihood, Discriminant power – are used in medical diagnosis. We show that they are interrelated, and we apply them to a case study from the field of electronic negotiations. We also list other learning problems which may benefit from the application of these measures.

Hsu, Yu-Feng et al. (2006) [7]In this paper, recent advances in computer technology have made digital image tampering more and more common. In this paper, we propose an authentic vs. spliced image classification method making use of geometry invariants in a semi-automatic manner. For a given image, we identify suspicious splicing areas, compute the geometry invariants from the pixels within each region, and then estimate the camera response function (CRF) from these geometry invariants. The cross-fitting errors are fed into a statistical classifier. Experiments show a very promising accuracy, 87%, over a large data set of 363 natural and spliced images. To the best of our knowledge, this is the first work detecting image splicing by verifying camera characteristic consistency from a single-channel image

Shi, Yun Q., et al. (2007) [8]In this paper, image splicing detection is of fundamental importance in digital forensics and therefore has attracted increasing attention recently. In this paper, we propose a blind, passive, yet effective splicing detection approach based on a natural image model. This natural image model consists of statistical features extracted from the given test image as well as 2-D arrays generated by applying to the test images multi-size block discrete cosine transform (MBDCT). The statistical features include moments of characteristic functions of wavelet subbands and Markov transition probabilities of difference 2-D arrays. To evaluate the performance of our proposed model, we further present a concrete implementation of this model that has been designed for and applied to the Columbia Image Splicing Detection Evaluation Dataset. Our experimental works have demonstrated that this new splicing detection scheme outperforms the state of the art by a significant margin when applied to the above-mentioned dataset, indicating that the proposed approach possesses promising capability in splicing detection.

Zhang, Zhen et al. (2008) [9] In this paper, to implement image splicing detection a blind, passive and effective splicing detection scheme was proposed in this paper. The model was based on moment features extracted from the multi-size block discrete cosine transform (MBDCT) and

© 2015-19, IJARCS All Rights Reserved

some image quality metrics (IQMs) extracted from the given test image, which are sensitive to spliced image. This model can measure statistical differences between original image and spliced image. Experimental results demonstrate that this new splicing detection algorithm is effective and reliable; indicating that the proposed approach has a broad application prospect.

.

3.ESTIMATION OF DCT COFFICIENTS MODEL PARAMETER

This subsection proposes to estimate the parameters (α, β) from unquantized DCT coefficients using the ML technique. By definition, the ML estimates ($\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$) are defined as the solution that maximizes the likelihood function:

$$(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \arg\max_{(\alpha,\beta)} \sum_{i=1}^{N} \log f_U\,(u_i)$$

We propose to calculate the variance and the kurtosis coefficient of U. Based on the law of overall expectation, the variance of U may be written as

$$Var_U[U] = E_U[U^2] = E_{\sigma^2}E_{U|\sigma^2}[U^2|\sigma^2] = E_{\sigma^2}[E_{\sigma^2}] = \alpha\beta$$

Wherein EX and Var_X represents the mathematical expectation and variance with appreciate to a random variable X. In addition, the kurtosis of U is given by using

$$\frac{E_U[U^4]}{Var_U^2[U]} = \frac{E_{\sigma^2}[\sigma^4]}{\mathbb{E}_{\sigma^2}^2[\sigma^2]} = 3\frac{\alpha\beta^2(\alpha+1)}{\alpha^2\beta^2} = 3\left(1+\frac{1}{\alpha}\right)$$

Therefore, the MM estimates ($\hat{\alpha}_{MM}$, $\hat{\beta}_{MM}$) can be given by

$$\hat{\alpha}_{MM} = \frac{3}{\frac{m_4}{m_2^2}-3}, \qquad and\hat{\beta}_{MM} = \frac{m_2}{\hat{\alpha}_{MM}},$$

wherem_k denotes the k-th sample moment of U.

4.METHODOLOGY

4.1 Proposed Estimation Algorithm

In order to avoid that situation, we propose to set $\tilde{Q}$ as the histogram bin containing the most nonzero rounded coefficients v. Then we filter all possible candidates of the true quantization step as follows:

$$S = \{\tilde{q}, \tilde{q}|\tilde{Q} and \varrho_{\tilde{q}} \geq t\}$$

where t, 0 ≤ t ≤ 1 is an empirical threshold. This strategy proposes to filter all integer divisors of $\tilde{Q}$ whose the ratio $\varrho_{\tilde{q}}$ is larger than a threshold t. Empirically, the threshold t is set to t = 0:75 in this work. In other words, this strategy proposes to start from the value $\tilde{Q}$ and refine all integer divisors of $\tilde{Q}$ based

Next, we verify each possible candidate $\tilde{q}$ in the set S by relying on the DCT coefficient statistics in order to provide an optimal estimate. Given the vector of IQF coefficients at the

samesubband, $\tilde{c} = (\tilde{c}_i)$, 1 ≤ i ≤ $n_{\tilde{q}}$, the optimal candidate q^* is the value maximizing the likelihood function

$$q^* = \arg\max_{\tilde{q}\in\mathcal{S}} \mathcal{L}\left(\tilde{\mathbf{c}} \mid \alpha, \beta, \tilde{q}\right) = \arg\max_{\tilde{q}\in\mathcal{S}} \sum_{i=1}^{n_{\tilde{q}}} \log P_C(c_i \mid \alpha, \beta, \tilde{q}),$$

where PC is the pmf of the quantized DCT coefficient We can note that the parameters $(\alpha, \beta$) in (25) are unknown. In order to overcome this difficulty, we propose to estimate the parameters (α, β) from rounded forward coefficients v. Consequently, the likelihood function $\mathcal{L}$ now only depends on one unknown parameter $\tilde{q}$, which leads to that the optimal candidate q^* can be given by

$$q^* = \arg\max_{\tilde{q}\in\mathcal{S}} \sum_{i=1}^{n_{\tilde{q}}} \log P_C(c_i \mid \hat{\alpha}, \hat{\beta}, \tilde{q}).$$

Subsequently, the principle steps of the set of rules for quantization step estimation from a given image are in the followings:

- **Step 1:** Divide the photo into 8 × 8 blocks and dispose of all uniform blocks and saturated blocks. The block is uniform if the maximum fee of block pixels is same to minimum price. The block is saturated if the most value is 255 or the minimum price is zero. These two styles of blocks are excluded for subsequent processing.
- **Step 2:** carry out DCT and rounding operation on the final blocks, and arrange the rounded ahead DCT coefficients v into sixty four vectors of coefficients of length N inside the zig-zag order.
- **Step 3:** For every sub band k, 1≤ ok ≤ sixty four, extract the vector v_k, estimate the parameters (α_k, β_k) counting on the technique, and determine the set S_k
- **Step 4:** determine the top-quality quantization step estimateq_k^*.

4.2 Proposed change in DCT forgery detection

The prevailing paintings proposes a histogram-based totally method for quantization step estimation. The principle contributions of the paintings are the followings:

- In comparison to earlier histogram-primarily based strategies, this paintings establishes a mathematical analysis of quantization effect all through JPEG compression and decompression in an effort to justify the relation of neighborhood maxima of the variety of integer quantized ahead (IQF) coefficients with the real quantization step. This evaluation famous that the quantity of IQF coefficients may be exploited as intrinsic quantization fingerprint for the trouble of JPEG compression records. The IQF fingerprint-based totally approach can provide a set of applicants of the authentic quantization step.
- Rather than the usage of the Laplacian model, the work includes the state-of-the-art statistical model of DCT coefficients into the estimation algorithm in order to enhance the accuracy. The version is used to provide the superior estimate of quantization step from the set of candidates given through the above method.
- The proposed technique is carried out on massive photograph databases to spotlight its excessive accuracy. Moreover, the proposed technique can provide dependable estimation of quantization steps for DC coefficients while a few present techniques fail. Additionally, the excessive performance on the color snap shots shows the robustness of the proposed technique to shade noise introduced for the duration of

© 2015-19, IJARCS All Rights Reserved

JPEG compression pipeline. Closing but now not least, the proposed method may be carried out to different sensible forensic scenarios consisting of estimation of the secondary quantization table in a double-JPEG compressed photo saved in lossless format and JPEG compression identification.

TABLE 4.1 NOTATIONS

X	Uncompressed input image
Y	JPEG-decompressed output image
u	Unquantized DCT coefficient
c	Quantized DCT coefficient
d	De quantized DCT coefficient
u	Forward DCT coefficient
V	Rounded forward DCT coefficient
$\bar{c}$	Quantized forward DCT coefficient
Q	Quantization step
N	Number of DCT coefficients at the same subband
$\bar{N}$	Number of convenient forward DCT coefficients
N	Number of integer quantized forward DCT coefficients
£	Round-off error

5.RESULT AND ANALYSIS

Fig 5.1 Tampered Input image for Analysis

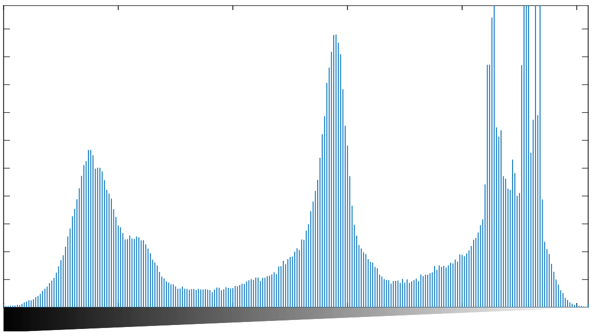

Fig 5.2 Histogram of R-Space (cr) of JPEG image Showing Sharp peaks indicator of forgery

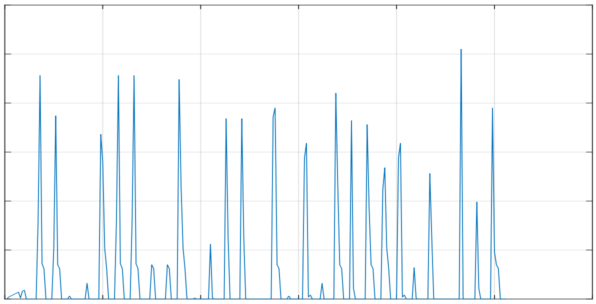

Fig 5.3 JPEG Huffman Symbols of input image

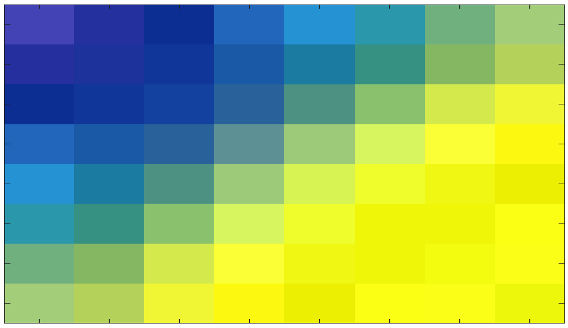

Fig 5.4 JPEG Quantization tables for Forgery Detection

Fig 5.5 Input image and Detected Regions

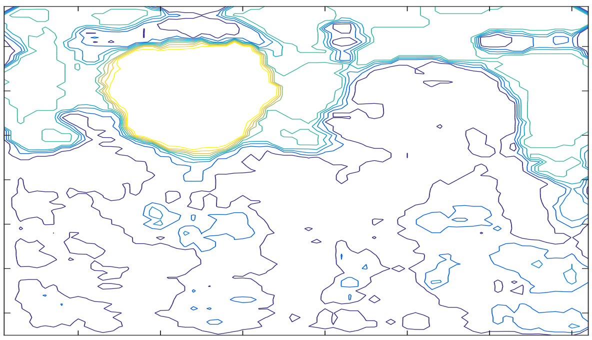

Fig 5.5 Contour of Detected Regions

Table 5.1 Overall Accuracy, TPR, TNR and AUC of Proposed Work

Forgery Type	Channel	Accuracy(%)	TPR(%)	TNR(%)	AUC

Copy-move	cr	97.6298803	98.24529	97.101804	97.1144
Splicing	cr	98.41792445	98.61898	97.353961	98.0568
Copy-move	cb	97.1635559	98.4027	97.62724	97.91724
Splicing	cb	98.96612488	97.23439	97.837317	97.35614
Copy-move	cr+cb	98.44034433	97.20708	97.106869	98.76725
Splicing	cr+cb	98.18587818	98.4454	98.114744	97.00798

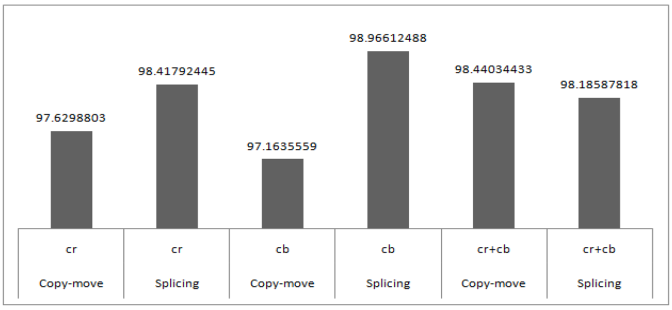

Fig 5.6 Accuracy of image forgery at varying channels

6. CONCLUSION AND FUTURE SCOPE

This paintings proposes an correct method for quantization step estimation from a given photo. By using reading the quantization impact in the course of JPEG compression and decompression pipeline, the paintings proposes a quantization fingerprint, specifically the number of integer quantized ahead coefficients, and provides a mathematical justification to reveal the relation of local maxima of that degree with the genuine quantization step. The estimation algorithm is designed by using relying on that quantization fingerprint and incorporating the statistical model of DCT coefficients that has been given in our previous works. Numerical experiments on big photograph database spotlight the relevance of the proposed technique. The electricity of the proposed method is the high estimation accuracy for a extensive type of snap shots with one-of-a-kind photograph contents, photograph sizes, and satisfactory elements. Moreover, the proposed approach can offer reliable estimation of quantization steps for DC coefficients while a few present strategies fail. The accuracy of the proposed method is emphasised while applying on actual color JPEG snap shots obtained from different digital camera models/brands. The excessive overall performance on the shade pictures shows the robustness of the proposed method to coloration noise introduced all through JPEG compression pipeline. In spite of its high performance in phrases of estimation accuracy, the proposed technique is greater time-eating than earlier-artwork ones. This disadvantage is because of the reality that the ML estimation of DCT model parameters is carried out the usage of a numerical optimization method. The proposed technique also indicates its accuracy in other practical forensic situations which includes estimation of the secondary quantization table in a double-JPEG compressed photo stored in lossless format, and JPEG compression identification.

7. REFERENCES

[1] Lee, Barry B., Joel Pokorny, Paul R. Martin, Arne Valbergt, and Vivianne C. Smith. "Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers." JOSA A 7, no. 12 (1990): 2223-2236.

[2] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20, no. 3 (1995): 273-297.

[3] Zhang, Guangcheng, Xiangsheng Huang, Stan Z. Li, Yangsheng Wang, and Xihong Wu. "Boosting local binary pattern (LBP)-based face recognition." In Advances in biometric person authentication, pp. 179-186. Springer Berlin Heidelberg, 2004.

[4] Ng, Tian-Tsong, Shih-Fu Chang, and Q. Sun. "A data set of authentic and spliced image blocks." Columbia University, ADVENT Technical Report (2004): 203-2004.

[5] Johnson, Micah K., and HanyFarid. "Exposing digital forgeries through chromatic aberration." In Proceedings of the 8th workshop on Multimedia and security, pp. 48-55. ACM, 2006.

[6] Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation." In Australasian Joint Conference on Artificial Intelligence, pp. 1015-1021. Springer Berlin Heidelberg, 2006.

[7] Hsu, Yu-Feng, and Shih-Fu Chang. "Detecting image splicing using geometry invariants and camera characteristics consistency." In Multimedia and Expo, 2006 IEEE International Conference on, pp. 549-552. IEEE, 2006.

[8] Shi, Yun Q., Chunhua Chen, and Wen Chen. "A natural image model approach to splicing detection." In Proceedings of the 9th workshop on Multimedia & security, pp. 51-62. ACM, 2007.

[9] Zhang, Zhen, Jiquan Kang, and Yuan Ren. "An effective algorithm of image splicing detection." In Computer Science and Software Engineering, 2008 International Conference on, vol. 1, pp. 1035-1039. IEEE, 2008.

[10] Dong, Jing, Wei Wang, Tieniu Tan, and Yun Q. Shi. "Run-length and edge statistics based approach for image splicing detection." In International Workshop on Digital Watermarking, pp. 76-87. Springer Berlin Heidelberg, 2008.

[11] Farid, Hany. "Image forgery detection." IEEE Signal processing magazine 26, no. 2 (2009): 16-25.

[12] Mahdian, Babak, and StanislavSaic. "A bibliography on blind methods for identifying image forgery." Signal Processing: Image Communication 25, no. 6 (2010): 389-399.

[13] Chih-weiHsu ,Chih-chung Chang , Chih-jen Lin, A practical guide to support vector classification, (2010).

[14] Wang, Wei, Jing Dong, and Tieniu Tan. "Image tampering detection based on stationary distribution of Markov chain." In Image Processing (ICIP), 2010 17th IEEE International Conference on, pp. 2101-2104. IEEE, 2010.

[15] Zhao, Xudong, Jianhua Li, Shenghong Li, and Shilin Wang. "Detecting digital image splicing in chroma spaces." In International Workshop on Digital Watermarking, pp. 12-22. Springer Berlin Heidelberg, 2010.

[16] Shivakumar, B. L., and Lt Dr S. SanthoshBaboo. "Detecting copy-move forgery in digital images: a survey and analysis of current methods." Global Journal of Computer Science and Technology 10, no. 7 (2010).

[17] Huang, Di, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. "Local binary patterns and its application to facial image analysis: a survey." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41, no. 6 (2011): 765-781.

[18] Hussain, Muhammad, SummrinaKanwalWajid, Ali Elzaart, and Mohammed Berbar. "A comparison of SVM kernel functions for breast cancer detection." In Computer Graphics, Imaging and Visualization (CGIV), 2011 Eighth International Conference on, pp. 145-150. IEEE, 2011.

© 2015-19, IJARCS All Rights Reserved

[19] Muhammad, Ghulam, Muhammad Hussain, and George Bebis. "Passive copy move image forgery detection using undecimated dyadic wavelet transform." Digital Investigation 9, no. 1 (2012): 49-57.

[20] Zhang, Yujin, Chenglin Zhao, Yiming Pi, and Shenghong Li. "Revealing image splicing forgery using local binary patterns of DCT coefficients." In Communications, Signal Processing, and Systems, pp. 181-189. Springer New York, 2012.

© 2015-19, IJARCS All Rights Reserved