International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-3, Issue-12, Dec.-2015

DIGITAL IMAGE FORENSICS USING NON-LINEAR DIMENSIONALITY REDUCTION TECHNIQUE

[1]ANIL DADA WARBHE, [2]R. V. DHARASKAR, [3]V. M. THAKARE

[1]Research Scholar, Department of Computer Science & Engineering, SGBAU, Amravati, MS, 444602
[2]Dr. R. V. Dharaskar, Former Director DES (Disha – DIMAT) Group of Institutes, Raipur, CG
[3]HOD, Post Graduate Dept. of Computer Science, SGB Amravati University, Amravati, MS, 444602
E-mail: [1]mtech2008@rediffmail.com, [2]rajiv.dharaskar@gmail.com, [3]vilthakare@yahoo.co.in

Abstract- Powerful and sophisticated digital image editing tools has made producing good quality forgeries very easily for almost everyone. Therefore, it is very important to identify the given image as original and without any modification. Digital Image Forensics is spurring field among researchers to answer this. In this context, proving the authenticity and integrity of digital images becomes increasingly important. In this paper, we propose a simple method to detect image tampering operations of the copy-paste forgery nature. We use a block based approach for the detection of copy-paste tampering in digital images. The proposed method uses nonlinear dimensionality reduction algorithm as a basis. Our preliminary results show that the proposed algorithm performs well for copy-paste forgery detection and can be used to identify tampered areas of the digital images.

Keywords- Digital Image Forensics, Copy-Paste Forgery, Dimensionality Reduction.

I. INTRODUCTION

With the availability of low cost off the shelf image manipulation and cloning tools, it is very easy to tamper and create fake images even to a person with lukewarm skills of photography. Every now and then we are been presented with the amazing and sometimes unbelievable kind of images in our email inboxes. Maximum of it are nothing but artificially synchronized photographic fakes, adopted, for promoting and floating different stories through media, emails and social networking websites. Figure 1 is a good example of such copy-paste forgery.

The manipulations done in the digital images cannot be detected and make out by naked eyes; as manipulations may not leave obvious evidence of tampering. The manipulation to change the original content of the image is also known as image fakery. The incidents of image fakery has grown tremendously during last few years due to the increasing usage of the social networking sites such as Facebook and message chats Whats App etc. Image fakery is a cybercrime, but because of the lack of proper regulatory framework and infrastructure for prosecution of such evolving cybercrime, leads to dissatisfaction about increasing use of such tools. This scene developed the feeling of cynicism and mistrust among civilians.

Similar to Cyber forensics [11], Mobile Forensics [12], over the past few years, the field of digital image forensics has emerged to help restore some trust to digital images. Hence, to detect such modifications in the original content in the images needs to be detected, and hence, the necessity of algorithms for efficiently verifying the integrity and authenticity of digital images cannot be overemphasized in this digital era.

Figure 1 (a) is original photo taken by Tien Chiu at Snake farm in Bangkok in year 2003. And (b) A Forged version of (a), an email hoax around 2010 claim that set of attached images depict a "rare" five headed Cobra that was found at Kukke Subrahmanya a Hindu temple located about 105 km from Mangalore in the state of Karnataka, in southern India [1].

There are many ways to create digital forgeries such as coping the portion from the image and pasting it on the same image to hide or conceal the original meaning of an image called copy-paste forgery. Copying the region from one image and pasting it on other image which is called as image splicing or compositing. In this paper we propose and algorithm for copy-paste type of image forgery.

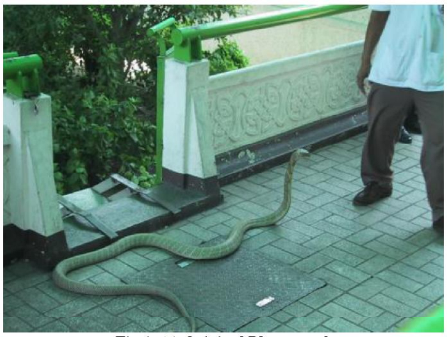

Fig.1. (a) Original Photograph

Fig.1. (b) Forged Photograph

II. RELATED WORK

The main aim of the image forgery detection especially that of a copy-move image forgery detection algorithm is to determine if a given image contains cloned regions without prior knowledge of their shape and location. An obvious approach is to exhaustively compare every possible pair of regions. However, such an approach is exponentially complex. Block matching appears to be a more efficient approach.

Utilizing such an approach, A copy-move image forgery detection algorithm is proposed in [2], and is based on PCA [3] which slides a b × b block over an N × N image pixel by pixel resulting in k = (N - b +1)2 blocks. Each block is column-wisely reshaped into a b2 long row vector, otherwise known as feature vector, and inserted into a k × b2 feature matrix. Principal Component Analysis (PCA) is performed to derive an alternative representation of each row of the feature matrix. PCA, which we present later in this section for the sake of completeness, is a well-known algebraic tool for matrix decomposition in literature. Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT) based method is proposed in [4], DWT is well explain in literature and widely used for its application to image processing [4], The proposed method in [5] uses the application of Principal Component Analysis-Eigenvalue Decomposition (PCA-EVD) in reducing the dimension of the feature vector while reducing the dimension of the image using DWT for the digital image forgery detection.

Zhao Junhong [6] presented a new technique based upon LLE (Locally Linear Embedding). LLE is method for dimensionality reduction in high dimensional data set suggested by Roweis [7]. LLE can find the topological relationship among nonlinear datasets and map high-dimensional data to low dimensional data without changing the relative locations. A digital image is essentially a nonlinear digital signal. The method works in a similar way as [2]. But for reduction of dimension LLE is being used instead of PCA. Because PCA lose some local information. LLE based method is better in detecting fused edges. The authors have compared the results of LLE based method with [2] and found that the copy moved area shown to be more pronounced. Recently M.K. Bashar et.al.[8] presented a method for detecting forgery in the presence of flip and rotation. The method has been applied with PCA, KPCA and wavelet transformed images. Different tables have been prepared for an exhaustive comparison between PCA, KPCA and wavelet based techniques for robustness against rotation, horizontal flips, vertical flips, translation and SNR. Also one method has been suggested for automation of threshold parameters for detecting duplicate regions. The authors claimed highest efficiency for KPCA based algorithm. However the time costs are high.

The algorithm proposed in this paper also uses KPCA. In next section a brief overview of KPCA is given.

III. KERNEL PCA

Natural images can be represented by principal component analysis (PCA). PCA attempts to efficiently represent data by finding orthonormal axes, which maximally de-correlate the data. However, PCA-based technique is useful if the spectral class structure of the transformed data is distributed along the first few axes. But the nonlinear correlations are not considered in the PCA case. Also PCA has a risk of losing information in the case of dimensionality reduction. A nonlinear projective approach like kernel principal component analysis (KPCA) can make up above deficiencies by adopting an appropriate kernel function and projecting the data to a relatively high dimensional space. .Kernel trick simplifies the process even more because there is no need to explicitly map the input vectors into the high dimensional feature spaces, all we need is to work with the dot products and dot products in the higher dimension feature space can be calculated in terms of dot product in input space i.e. lower dimension without explicitly calculating the mapping.

Some key points of KPCA:

- KPCA extracts more useful features than the linear PCA.
- Initial mapping to high-dimensional space provides us with smoother dimensionality reduction than standard PCA.
- It does not require nonlinear optimization but just the solution of eigenvalue problem.

Kernel Trick:- Given any algorithm that solely depends on the dot product of the input vectors, we can work by mapping input into a higher dimension feature space and work with dot products in this space which can be computed using dot products in low dimension space(input space) without having the need to calculate the mapping explicitly.

IV. METHOD

We use a block based approach for the detection of copy-paste tampering in the natural images. An unknown color image is first converted into its grayscale version, which is then divided into small overlapping blocks of n pixels. Two matrices are created, one called training and the other being test matrix. Now we transform the input data to a higher dimensional feature space implicitly using kernel trick. The kernel that we use here is a RBF kernel specifically Gaussian-RBF kernel.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

We assume the mapped data in higher dimensional feature space is centered. We can write the covariance matrix as

$$C = \frac{1}{M}\sum_{i=1}^{M} \Phi(x_i)\Phi^T(x_i)$$

We compute the training gram matrix Ktrn and test gram matrix Ktst using the Gaussian kernel. The centered versions of these gram matrices are:

$$\widetilde{K}_{trn} = K_{trn} - 1_M K_{trn} - K_{trn} 1_M + 1_M K_{trn} 1_M$$

$$\widetilde{K}_{tst} = K_{tst} - 1'_M K_{trn} - K_{tst} 1_M + 1_M K_{trn} 1_M$$

For derivations of above refer [8]. Now we compute eigenvalues and corresponding eigenvectors of the centered training gram matrix and normalize the eigenvectors. Now we sort the eigenvalues in decreasing order and also sort eigenvectors accordingly. Intuition is that we are ordering the eigenvectors such that there is maximum variance associated with the first eigenvector and variance reduces as we move along the decreasing sorted sequence of eigenvalues. The centered test matrix is projected along the ordered eigenvectors. The first component will have maximum variance. Next will be the second component and so on. Henceforth we will work with projected version of the centered test data matrix.

Now the detection of the copy-paste starts. We sort the rows of the projected version matrix (sorting lexicographically). Intuition being that all similar blocks will be nearby to each other so that they can be identified easily. Now we compute offsets between ith row and jth row (j>=I and j-i<=Rth) in the following way

$$(m,n)=\{(xi-xj,yi-yj) \quad if \quad (xi-xj)>0,(xj-xi,yj-yi) \quad if \quad (xi-xj)<0,(0, \quad |yi-yj|) \quad if \quad (xi=xj)$$

In the above equation, (m, n) are the x and y offsets corresponding to the ijth pair of blocks. The offset frequency is defined as the sum of all occurrences of an offset and may be given by

$$Km,n=\#(m,n)$$

We can construct a list of the offset frequencies and identify the duplicated regions by using the threshold on them. The basic idea is that duplicated regions consist of many such small blocks, which appear in close proximity in the lexicographically sorted matrix. Since the original copy and its duplicated regions appear in pair, the blocks representing such areas must have the same offsets. Therefore we have found the duplicated regions.

V. ALGORITHM

Here are the proposed algorithm steps to detect copy move forgery

1. Load the given image.
2. Convert the image given into grayscale.
3. Construct a training matrix from the grayscale image by blocking the image (every pixel need not be the start of a block).
4. Construct a testing matrix from the grayscale image by blocking the image (every pixel must be start of a block).
5. Construct a training gram matrix from the training matrix using Gaussian kernel.
6. Construct a testing gram matrix from training and testing matrix using Gaussian kernel.
7. Center the training gram matrix as described in the previous section.
8. Center the testing gram matrix as described in the previous section.
9. Find the eigenvectors and eigenvalues of the centered training gram matrix.
10. Sort the eigenvalues in decreasing order and also the eigenvectors accordingly.
11. Compute the projections of the centered testing gram matrix on the ordered Eigen vectors (decreasing order of eigenvalues).
12. Lexicographically sort the projected version of the centered testing gram matrix. Let it be denoted by S.
13. For every ith row si in S, select a number of subsequent rows, Sj such that |i-j|<=Rth and place all the pairs of coordinates (xi, yi) and (xj, yj) on to a list Pin.
14. Compute offset for each row of Pin.
15. Compute the frequency offset.
16. Those rows in Pin which have high frequency offsets are the duplicated regions.

VI. EXPERIMENTAL RESULTS

In our experiments, we have tested different sized forged images. Followings are the experimental results. The Dth Parameter has been needed to be set for the effective detection of copy-moved regions.

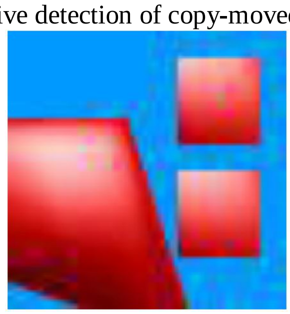

Fig.3. Forged Photograph

Fig.4. Detected Forged Regions

Fig.2. Tempered image of cherries

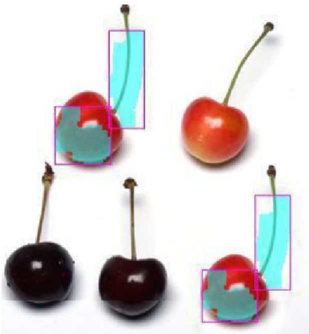

Fig.2. Detection of forged regions using proposed Method

CONCLUSIONS

In this paper, we have proposed a nonlinear dimensionality reduction algorithm which uses block based approach to detect Copy-Paste forgery in images automatically and effectively. The threshold parameter such as Dth is needed to be set. The algorithm works well if the regions from the same images are copy pasted, but do not identify the pasted regions if they are scaled and interpolated.

REFERENCES

[1] http://www.hoax-slayer.com/five-headed-cobra-hoax.shtml.

[2] A. C. Popescu and H. Farid, ―Exposing Digital Forgeries by Detecting Duplicated Image Regions,‖ Technical Report, TR2004-515, Department of Computer Science, Dartmouth College, 2004.

[3] R. Duda and P. Hart. Pattern Classification and Scene Analysis. John Wiley and Sons, 1973

[4] G. Li, Q. Wu, D. Tu, and S. Sun, ―A Sorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries based on DWT and SVD,‖ in Proceedings of IEEE International

[5] Conference on Multimedia and Expo, Beijing China, July 2-5, 2007, pp. 1750-1753.

[6] Michael Zimba, Sun Xingming , ―DWT-PCA (EVD) Based Copy-move Image Forgery Detection‖, International Journal of Digital Content Technology and its Applications. Volume 5, Number 1, January 2011

[7] Zhao Junhong ―Detection of Copy-Move Forgery Based on One Improved LLE Method‖, vol. 4, pp 547-550, 2nd International Conference on Advanced Computer Control (ICACC), 2010.

[8] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000. 290(5000): 2323 - 2326

[9] M.K. Bashar, K. Noda, N. Ohnishi, K. Mori , ‖Exploring Duplicated Regions in Natural Images‖, IEEE Transactions on Image Processing,2010.

[10] Anil D. Warbhe, R. V. Dharaskar, "Blind Method for Image Forgery Detection: A tool for Digital Image Forensics", International Journal of Computer Applications, 2012.

[11] Anil D. Warbhe, R. V. Dharaskar, "Survey on Pixel-Based and Format-Based Image Forgery Detection Techniques", National Conference MPGI, Nanded-2012.

[12] Nirkhi, S.M.; Dharaskar, R.V.; Thakre, V.M., "Analysis of online messages for identity tracing in cybercrime investigation," in Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on , vol., no., pp.300-305, 26-28 June 2012.

[13] Ahmed, Rizwan, and Rajiv V. Dharaskar. "Study of mobile botnets: An analysis from the perspective of efficient generalized forensics framework for mobile devices." IJCA Proceedings on National Conference on Innovative Paradigms in Engineering and Technology (NCIPET 2012) ncipet (15). 2012.

★★★