



Exploring Forgeries by Detecting Duplicated Image Regions

Ms.K.V.Hulmukhe and Mr.M.B.Jhade

Department of Computer Engineering,
KKWIEER,Nashik

Abstract: With the innovation in technology and availability of imaging tools, it's not difficult now days to manipulate digital images to hide or add content in images. Image forgery detection is currently one of the hot research fields of image processing.

Image forgery detection approach that can adopt two robust features based on discrete wavelet transform (DWT) and principal component analysis (PCA). KPCA(Kernel PrincipalComponent Anaysis) extracts non linear features that are useful than the linear PCA. KPCA represent image data into to high-dimensional space without reducing dimension than the standard PCA. KPCA schemes provide excellent representations of the image data for robust block matching. Experiments with a good number of images shows very promising results, when compared with the conventional DWT,PCA and KPCA based- approach.

Keywords: DWT;;forgery detection;KPCA;Linear transform;Non-linear transform;PCA

I. Introduction

Digital images are easy to tamper because of the availability of the powerful editing software and sophisticated digital cameras. Some software's like Photoshop, 3DS Max, photoscape etc. are so sophisticated that it is very hard to differentiate tampered images. From the tabloid magazines to the fashion industry and in mainstream media outlets, scientific journals, political campaigns, courtrooms and the photo hoaxes that land in our e-mail in-boxes, doctored photographs are appearing with a growing frequency and sophistication. As a result, digital evidences have not yet been accepted in real life applications, for example, criminal investigation, medical reporting etc.

Hence the need for some technological solutions to determine if an image has been tampered or not.

One of the common cases of tampering an image is the copy-paste tampering. For example, sometimes objects in the real photograph are cut and paste on some part of image.

One solution to detect duplicated tampered region from an image is to apply the watermarking to the image. But adding watermarking feature in cameras is an additional cost. Most cameras do not have this feature. So this method cannot be used in general.

The above problem is solve by implementing an algorithm to detect the copy-paste or region-duplication in images using DWT, PCA & extension of PCA is KPCA block based methods are used.



Fig1: Example of copy-move forgery

In the above image, the image on the left side is the real image whereas the one on the right is the forge image of the left one. A portion of the image on the left image is used to cover by the dog in the image. In this way, multiple image objects can be added ,hide or deleted from an image.

Exploring forgery detection for digital images is, therefore, an important research area.

This system will focus on not only copy move type of image forgery but also for detecting different geometric operation using DWT,PCA and to improve these method of forgery detection extension of PCA i.e KPCA approach is used which provide excellent representations of the image data for robust block matching.

To implement an algorithm for detecting duplicated image region in digital images, different approaches are used in this system for representing the image data which provide robust image feature for detecting duplicated image region

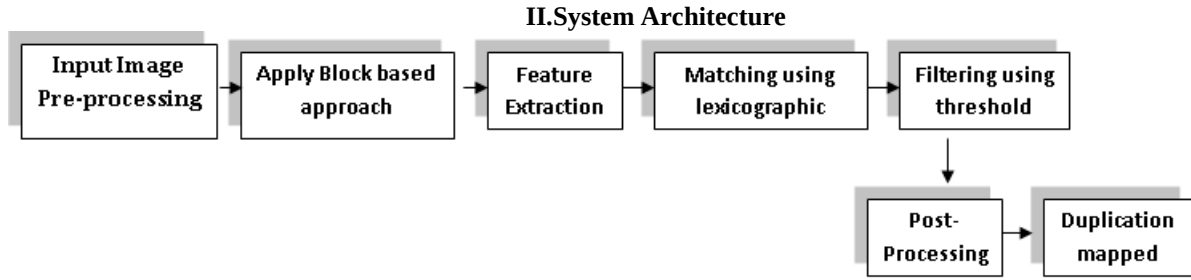


Fig.2 Processing Steps for Detecting Duplicated Image Region

A. Description of the method used

- DWT

Wavelet coefficients are extracted from the sub bands after discrete wavelet transform. First applying multi-resolution wavelet decomposition to small fixed-sized image blocks. This transform decomposes an image with an overall scale factor of 4 providing, at each level, one low-resolution sub-image (LL) and three wavelet Coefficients sub-images (HL, LH, HH)[1].



Fig.3

(a)Original Image (b)Wavelet sub-bands(LL,HL,LH,HH)r representation after one level decomposition

For image forgery detection 2 level wavelet decomposition is used for calculating wavelet coefficients for each block. These wavelet coefficients are then stacked into a vector for each block. These feature vectors are used for further processing

- PCA

1. Let $\vec{x}_i, i = 1, \dots, N_b$ denote these blocks in vectorized form

2. Computer the covariance matrix C as the eigen vectors $\vec{e}_j$ of the matrix C , with corresponding eigenvalues as

$$C \vec{e}_j = \lambda_j \vec{e}_j \quad (1)$$

3. To define the principal components, where $j = 1, \dots, b$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_b$. The eigenvectors, $\vec{e}_j$ form a non linear basis for each image block, $\vec{x}_i$

$$\vec{x}_i = \sum_{j=1}^b a_j \vec{e}_j \quad (2)$$

$$\text{Where } a_j = \vec{x}_i^T \vec{e}_j \quad (3)$$

The dimensionality of this representation can be reduced by simply truncating the sum in Equation (3) to the first N_t terms. Note that the projection onto the first N_t eigenvectors of the PCA basis gives the best N_t -dimensional approximation in the least squares sense[13].

These PCA components of each block are used for further processing for detecting duplicated image region.

- KPCA

Kernel PCA is a powerful nonlinear feature detector, which has been suggested for various image processing tasks such as compression and denoising requiring. It has been successfully applied to the representation of the natural image. So KPCA is use for detecting tampered regions in the natural images.

Some key points of KPCA

1. KPCA extracts more useful features than the linear PCA[1].

2. Initial mapping to high-dimensional space provides us with smoother dimensionality reduction than standard PCA[1].

3. It does not require nonlinear optimization but just the solution of eigenvalue problem[1]

By using a nonlinear kernel function instead of the standard dot product, we implicitly perform PCA in a possibly high dimensional space which is nonlinearly related to input space, where eigenvectors form a new

nonlinear basis. Implimentation of KPCS consists of three different layers, namely input, hidden and output layers. The input layer is made up of source nodes. The input is Test pattern x_t , comes from an 16×16 block in the input image. As per image tampering problem, input to a kernel PCA are all pixel intensities corresponding to a data-block[1].

The hidden layer computes a dot product and map input space to the feature space F . These two operations are performed in one single step using the kernel function $K(x_t, x_i) = \Phi(x_t) \cdot \Phi(x_i)$ [1]. The outputs of the hidden layer are then linearly combined using weights α_i , resulting in a nonlinear principal component corresponding to Φ . selected Gaussian kernel function, kernel hyper parameter σ , training samples M and the block-size b , KPCA can be implemented [1].

III. Duplicate Detection Algorithm

1. Provide input image of size $M \times N$. Convert the image to gray scale image
2. Tile the image in N_b overlapping blocks of size $b \times b$. Each block having fixed size 16×16
Where $0 \leq i < ((M - b + 1) \cdot (N - b + 1))$
3. Use 2D wavelet transform in case of DWT transform decomposes an image with an overall scale factor of 4 providing, at each level, one low-resolution sub-image (LL) and three wavelet coefficient sub-images (HL, LH, HH)[1] and Using PCA, compute the new N_t -dimensional representation, $\vec{a}_i$, $i = 1, \dots, N_b$, of each b pixel image block The value of N_t is chosen to satisfy: $1 - \varepsilon = \frac{\sum_{i=1}^{N_t} \lambda_i}{\sum_{i=1}^b \lambda_i}$ [13] where λ_i are the eigenvalues as computed by the PCA or KPCA.
4. Select Number of Quantization bins Q . Construct a data matrix, M_{data} , where row-elements contain component-wise quantized features, i.e., $\frac{\vec{a}_i}{Q}$.
5. Apply lexicographic sorting to the rows of M_{data} matrix to obtain a new matrix S . Let s_i , be the i th row of S , which represents the i th block with its center coordinates (x_i, y_i) .
6. Select row threshold parameter R_{th} R_{th} is used to select number of adjacent rows to search.
For every pair of rows $\vec{s}_i$ and $\vec{s}_j$ from S such that $|i - j| < R_{th}$ place the pair of coordinates (x_i, y_i) and (x_j, y_j) onto a list F_{in} .
7. D_{th} minimum offset magnitude threshold Discard all pairs whose offset magnitude,
 $Dof = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} < D_{th}$. Create a refined list of point pair F_{out} from F_{in}
7. From the remaining pairs of blocks in F_{out} label duplicated region by constructing a zero image of the same size as the original, and coloring all pixels in a duplicated region with a unique grayscale intensity value[1].

IV. Experiments

In this system assume that all Test images are taken by digital cameras

Experimental observations showed that KPCA-based feature produces larger average precisions as well as recalls, when compared with the wavelet- or PCA-based feature in case of noisy or compressed images with optimized parameter settings. The reason is due to representing image data using and lexicographic sorting on them[1].

KPCA based feature produces better performance result as compared to the wavelet- and compressed environments.

A. Results of Duplicate Detection using DWT



Fig.5

(a)Original Image (b)Input Image (c)Output Image

Result Analysis after applying DWT for test image1

Image size: 157×128

Number of Blocks: 55,460

Rth: 50

Elapsed time : 1645.879334 seconds.

B. Results of Duplicate Detection using PCA



Fig.8

(a)Original Image (b) Input Image (c) Output Image

C. Result of Duplicate Detection using KPCA



Fig.11

(a)Original Image (b)Input Image (c)Output Image

Result Analysis after applying KPCA for test image1

Image size:157×128

Number of Blocks:55,460

R_{th} :50

Elapsed time : 6438.497598 seconds.

It is clear from the results of the experiment that DWT,PCA and KPCA are block based methods are used for detecting the copy-paste or region-duplication in natural image.KPCA (Kernel Principal Component Anaysis) extracts more useful features than the linear PCA and also solve dimentionality reduction problem.

V. Conclusion

The aim behind "Exploring Forgery by Detecting Duplicated Image Region" is implying to detect duplicated image region from the digital image. This system is mainly based on detecting tampered data in form of duplicated image region .It uses variousrobust block-based feature approaches such as DWT,linear transformer PCA and Non linear transformer Kernel PCA. The results generated after comparing different blockbased approaches for detecting duplicated image region, KPCA(Kernel Principal Component Anaysis) extracts more useful features than the linear PCA. Initial mapping to high-dimensional space provides us with smoother dimensionality reduction than the standard PCA.

KPCA schemes provide excellent representations of the image data for robust block matching. Hence, this system is suitable for applications in which digital evidences in the form of content tampering have not been accepted.

Other heuristics or approaches for estimating the kernel σ parameter may be considered in future. Another challenge is, for detecting forgery in case of simultaneous multiple geometricoperations.

VI. References

- [1] M. K. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring Duplicated Regions in Natural Images". 2010 IEEE.
- [2] Hany Farid, "Image Forgery Detection", IEEE Signal processing magazine, March 2009.
- [3] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions," Technical Report, TR2004-515, Dartmouth College, Computer Science, 2004.
- [4] H. Farid, "Digital Doctoring: How to tell the real from the fake,"Significance, vol. 3, no. 4, pp. 162-166, 2006.
- [5] C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions," Technical Report, TR2004-515, Dartmouth College, Computer Science, 2004.
- [6] J. Fridrich, D. Soukal, and J. Lukas, "Detection of copy-move forgery in digital images," in Proc. Digital Forensic Workshop, August, 2003.

- [7] B. Scholkopf, A. Smola, and K. -R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [8] A. N. Myna, M. G. Venkateshmurthy, and C. G. Patil, "Detection of Region Duplication Forgery in Digital Images Using Wavelets and Log- Polar Mapping," in *Proc. Int. Conf. Comput. Intell. Mul. App.*, vol. 03, pp. 371-377, 2007.
- [9] G. Li, Q. Wu, Dan Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on dwt and svd," in *Proc. ICME 2007*, pp.1750-1753, July 2-5, 2007.
- [10] H. Huang, W. Guo, and Y. Zhang, "Detection of Copy Move Forgery in Digital Images Using SIFT Algorithm," in *Proc. IEEE Workshop Comput.Intell. Ind. App.*, vol.2 pp.272-276, 2008.
- [11] "Learning with kernels" by Scholkopf and Smola.
- [12] "Kernel Principal Component Analysis" Bernhard Scholkopf, Alexander Smola
- [13] Exploring Digital Forgeries by Detecting Duplicated Image Regions AlinCPopescu and HanyFarid,
Department of Computer Science Dartmouth College HanoverNH03755

VII.Acknowledgments

I would like to express deep sense of gratitude to my respected guide Prof.M.B.Jhade for their valuable comments during the progress of this paper. Lastly I would like to thank all our lectures and staff members of the Computer Department of K.K.W.IEER Nasik who helped me directly or indirectly.