Suitability of Independent Component Analysis in Digital Image Forgery Detection

Sunil Kumar [#1], J.V. Desai [#2], S. Mukherjee [*3], P. K. Das [#]
[#] Faculty of Engineering & Technology
Mody Institute of Technology & Science
Lakshmangarh, India
[*] Department of Electrical Engineering
Indian Institute of Technology
Roorkee, India
[1] skvasistha@ieee.org
[2] jagandesai@yahoo.com
[3] mukherjee.shaktidev@gmail.com

***Abstract*— Digital image forgery detection is one of the hot research area in the recent time. A lot of researchers are trying different tools to establish the authenticity of a given image. There can be many types of forgery performed on digital image. Watermarking is one of the traditional techniques to detect any type of tampering with the original image, but that has to be done at the time of capturing the image. Once an image has been captured without such technique, there is no alternate but to apply different blind forgery detection techniques. The present paper is an effort to explore Independent Component Analysis (ICA) as a tool to get clues about the tampering with original image. The results may provide further leads to the researchers working in the same area.**

Keyword- digital image forgery, ICA, image tampering, blind source separation

I. Introduction

Digital image forgery can be performed in today's world very easily with the kind of software tools and ever improving hardware. It can be performed in different ways [1]. It may be as simple as inserting some object from other image to elude the viewer or as complicated as pasting part of the same image to hide some information after various translations and rotations on the cropped part. Broadly severe type of forgery can be classified in to splicing and copy-move attack. Recently there has been more emphasis on copy move attack and comparison and effectiveness of the popular methods have been reported in surveys like [2] and [3]. In spite of the many existing methods there is lack of inclusive approach to deal all kind of image tampering. In the noted trends of forgery more than one image are mixed and the resultant forged image is created. So essentially, if it can be established that the resultant image is a mixture of multiple images, then it can be termed as a forged image. Independent component analysis (ICA) is well known for its ability to separate the sources, in case the sources are statistically independent. As ICA has the ability to extract the independent components in mixed audio signal as in cocktail party problem [4], it can well be used for independent component extraction in images also. ICA is already being used in the areas of pattern recognition and medical imaging, but it has not been tried extensively to detect tampering in digital images.

II. Independent component analysis

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed nongaussian and mutually independent and they are called independent components of the observed data. These independent components, also called sources or factors, can be found by ICA. The ICA algorithm was initially proposed in [5] to solve the blind source separation (BSS) problem i.e. given only mixtures of a set of underlying sources, the task is to separate the mixed signals and retrieve the original sources. Neither the mixing process nor the distribution of sources is known in the process. A simple mathematical representation of the ICA model [6] is as follows:

Consider a simple linear model which consists of N sources of T samples i.e. $Si = [Si(1), ...,Si(t),...,Si(T)]$. The symbol 't' represents time, but it may represent some other parameter like space. M weighted mixtures of the sources are observed as X, where $Xi = [Xi(1),... ,Xi(t),... ,Xi(T)]$. This can be represented as: $X = A S + n$

Where $X= (X_1, X_2, X_3………., X_M)$; $S= (S_1, S_2, S_3…., S_N)$ and $n= (n_1, n_2, n_3………., n_k)$ where n represent the additive white Gaussian noise (AWGN). It is assumed that there are at least as many observations as sources i.e. M = N. The M × N matrix A is represented as

$$A=\begin{bmatrix} a_{11} & . & . & . & a_{1N} \\ a_{21} & . & . & . & a_{2N} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{M1} & . & . & . & a_{MN} \end{bmatrix}$$

A relates X and S. A is called the mixing matrix. The estimation of the matrix S with knowledge of X is the linear source separation problem. If the additive noise n is negligible, then the original sources can be estimated by evaluating the pseudo inverse of the matrix A, which is known as the un-mixing matrix B, such that BX = BAS = S. For the cases where the number of observations M equals the number of sources N (i.e. M = N), the mixing matrix A is a square matrix with full rank and $B = A^{-1}$. The necessary and sufficient condition for the pseudo-inverse of A to exist is that it should be of full rank. When there are more observations than the sources (i.e. M > N), there exist many matrices B which satisfy the condition BA = I. Here the choice B depends on the components of S that we are interested in. When the number of observations is less than the number of sources (i.e. M < N), a solution does not exist, unless further assumptions are made. If there is no prior knowledge of the mixing matrix A, then the estimation of both A and S is known as a blind source separation (BSS) problem. Estimation of the underlying independent sources is the primary objective of the BSS problem. The problem is solvable under the assumption of negligible Gaussian noise n, with the following restrictions:

1. The sources (i.e. the components of S) are statistically independent.
2. At most, one of the sources is Gaussian distributed.
3. The mixing matrix is of full rank.

Before applying ICA whitening is performed on the observed vector X. This means the observed vector is transformed linearly so that we obtain a new vector which is white, i.e. its components are uncorrelated and their variances equal unity. In other words, the covariance matrix of transformed variable X' equals the identity matrix: $E\{X'X'^T\}=I$. One method for whitening is to use eigen value decomposition (EVD) of the covariance matrix $E\{XX^T\}=EDE^T$, where E is orthogonal matrix of the eigen vector $E\{XX^T\}$ and D is diagonal matrix of its eigen values. $E\{XX^T\}$ can be estimated in a standard way from the available sample X(1)…X(M) . Whitening can now be done by $ED^{-1/2}$ $X'=ED^{-1/2}E^TX$, where $D^{-1/2}=\mathrm{diag}(d_1^{-1/2}, d_2^{-1/2}, \ldots, d_n^{-1/2})$.

III. The method

In the present work Fixed Point ICA [7] is applied on different images. FPICA works on the principle of minimization and maximization of Kurtosis. It converges faster than the other ICA variations. Two forged images are taken as the mixed signals and provided as inputs to FPICA. The estimated independent sources are given as output. The method has been applied to detect three kinds of tampering. In case one, copy move attack is used for creating the mixed (forged) images and in case two, splicing has been used to forge the image and in the last case rotation and translation has been used to forge the image.

IV. Results

The FPICA algorithm is applied on three cases for different image sets. Two mixed images are given as input and two independent components are extracted. The outputs are shown case by case. In the first case two forged images in fig 1(b) and 1(c) are created by using copy move attack on original image fig 1(a). One statue has been copied and pasted at two different locations. Fig 1(d) shows the output of FPICA as independent components. The copied part has been separated as independent component. Also it shows that the same part is copied in both forged images at different places. Fig 2(b) and 2(c) are showing the forged images created by splicing a bird image to the original image of fountain in fig 2(a). Fig. 2(d) shows the results of independent component separation. In the third case spliced images in fig 3(b), fig 3(c) are created by scaling and tilting a book photograph and then morphing with original image in fig. 3(a).The sizes of the pasted regions are different and there is difference of 20 degrees in tilt . Fig. 3(d) shows the results of the third case. The time taken and number of iterations for source separation for all three cases is shown in table-1.

Case No.	Size of forged image (pixels)	No of trials and iterations for extracting independent component1	No of trials and iterations for extracting independent component2	Time taken by FPICA
1.	774x518	3,11	1, 11	147.34 seconds
2.	653x490	1,11	1,11	1.19 seconds
3.	653x490	1,11	1,11	1.33 seconds

Table-1 (showing performance of FPICA for the three cases)

Case1: Mixed images are generated by copy-move attack

Figure-1(a): Original image having four statues

Figure-1(b): First forged image created by copy and pasting one of the statues

Figure-1(c): Second forged image created by copy pasting one statue at different place

Figure-1(d): Estimated sources by applying blind source separation on the forged images 1(b) and 1(c). Output1 is showing the copy pasted region and output2 showing original image with the locations of pasted region

Case 2: Mixed images are generated by splicing

Figure-2(a): Original image

Figure-2(b): First forged image created by pasting external image of the bird on top left corner

Figure-2(c): Second forged image is created by pasting external image of the bird on the top right corner

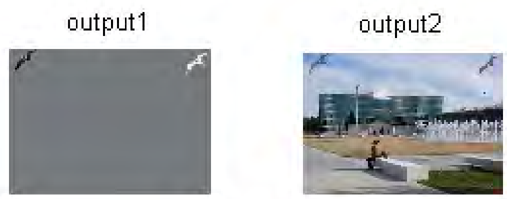

Figure-2(d): Estimated sources by applying blind source separation on the forged images 2(b) and 2(c). Output1 is showing the copy pasted region and output2 showing original image with the locations of pasted region

Case3: Mixed images are generated by scaling and tilting


Figure-3(a): Original image


Figure-3(b): First forged image is created by pasting a book image on the bottom right corner


Figure-3(c): Second forged image is created by scaling the pasted region as well as tilting by 20 degrees

Figure-3(d): Estimated sources by applying blind source separation on the forged images 3(b) and 3(c). Output1 is showing the copy pasted region and output2 showing original image with the locations of pasted region

V. CONCLUSION

The results have shown the promise of ICA for blind source separation even in case of images. This can be used to establish the truth of digital images particularly, if more than one version is available to act as mixed signals. Even if one of them is the original, then also spliced or copy moved region should be shown as independent component. Also from the third case it's evident that ICA is capable to handle the manipulations like scaling and tilting, which the other methods find difficult to handle efficiently. However in the presented experiment limited numbers of images are tested and the experiment may be conducted for more diverse image datasets. Also at least two variations of equal size of the original images are required to apply this method. ICA may be further explored for its ability even in the cases where only one image is available.

REFERENCES

[1] Farid, H. (2004). Creating and detecting doctored and virtual images: Implications to the child pornography prevention act. Department of Computer Science, Dartmouth College, TR2004-518, 13.
[2] Christlein, Vincent, Christian Riess, Johannes Jordan, and E. Angelopoulou. "An Evaluation of Popular Copy-Move Forgery Detection Approaches."IEEE Transactions on Information Forensics and Security (2012): 1-1.
[3] Kumar, S., and P. K. Das. "Copy-Move Forgery Detection in Digital Images: Progress and Challenges." International Journal on Computer Science and Engineering 3.2 (2011): 652-663.
[4] Haykin, Simon, and Zhe Chen. "The cocktail party problem." Neural computation 17.9 (2005): 1875-1902.
[5] P. Comon, "Independent Component Analysis-A new concept?" Signal Processing, vol. 36, pp. 287-314, 1994.
[6] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." Neural networks 13.4 (2000): 411-430.
[7] Hyvarinen, Aapo. "Fast and robust fixed-point algorithms for independent component analysis." IEEE Transactions on Neural Networks 10.3 (1999): 626-634.