Steganalysis: Using the blind deconvolution to retrieve the hidden data

P. Jidesh and Santhosh George
Department of Mathematical and Computational Sciences,
National Institute of Technology, Karnataka-575025,
India.
e-mail: jidesh@nitk.ac.in, sgeorge@nitk.ac.in

***Abstract*— Steganography has gained a substantial attention due to its application in wide areas. Steganography as it literally mean is hiding the information (stego data) inside the data (communication data) so that the receiver can only extract the desired information from the data. *Steganalysis* is the reverse process of steganography in which the information about the original data is hardly available, from the received data the extractor needs to identify the original data. Since this belong to a class of inverse problems it is hard to find the approximate match of the original data from the received one. In most of the cases this will fall under the category of ill-posed problems. The stego-data that has been embedded into the communication data can be considered as linear bounded operator operating on the input data and the reverse process (the Steganalysis) can be thought like a deconvolution problem by which we can extract the original data. Here we are assuming the watermarking as a linear operation with a bounded linear operator $K : X \to Y$ where X and Y are spaces of Bounded Variation (BV). The forward problem (the Steganography) is a direct convolution and the reverse (backward) problem (*steganalysis*) is a de-convolution procedure. In this work we are embedding a Gaussian random variable array with zero mean and with a specific variance into the data and we show how the original data can be extracted using the regularization method. The results are shown to substantiate the ability of the method to perform *steganalysis*.**

Keywords- Steganography, Steganalysis, Regularization, Inverse Problems, Image Processing.

I. INTRODUCTION

The idea of embedding data into the images has survived for many decades, the embedding technique which is widely known as digital watermarking took its shape a couple of decade ago [1]. Thereafter there was a tremendous growth in the techniques used for embedding the watermark. The watermark was embedded in the spatial domain of the input image, either visibly or invisibly [2], apart from the spatial domain, the data was embedded in transform domain as well [3], in order to exploit the detailed information available in the transformed domain. Although many techniques are available in literature for watermarking, most of them fail to extract the watermark without prior knowledge of the original data on which the watermark is embedded or the watermark itself. The process of retrieving the watermark without the prior knowledge of the input data is called the s*teganalysis* [4].

978-1-4244-8679-3/11/$26.00 ©2011 IEEE

In this paper we consider the watermark as a kernel which is embedded in the image with a linear operator (convolution). This linear operation can be modeled using Fredholm integral equation as:

$$g(s) = \int k(s,t)u(t)dt \qquad (1)$$

where $k(s,t)$ is the kernel which represents the watermark, $u(t)$ denotes the original signal that needs to be recovered from the observed one $g(s)$. The eq. (1) is a general equation for one dimensional continuous function. Since we are dealing with images, we can modify the above equation to:

$$\int\int_{\Omega} k(s,t;s_0,t_0)u(s,t)\,ds\,dt = z(s_0,t_0) \qquad (2)$$

here Ω denotes the image domain where the watermark has to be applied. Since we adopt a global watermarking strategy, Ω represents the whole image domain. Whereas $u(s,t)$ is the original input image we need to find from the observed image or watermarked image $z(s_0,t_0)$. The discretization of (2) with quadrature method will end up in a standard form $Ax = b$, here A is a bounded linear operator from $X \to Y$ where X and Y are the spaces of Bounded Variations (*BV*), x is the input image and b is the observed image. This problem is an inverse problem in which we have to find the input data x from the observed data b with less or minimum knowledge about the data embedded into this. So the trivial solution can be written as $x = A^{-1}b$ where A^{-1} is the inverse operator, in our case it is an inverse convolution operator (deconvolution). For simplicity, we assume that the linear operator is a 2-D convolution operator and the kernel is a Gaussian kernel with mean μ =0 and variance σ^2.

Now the trivial solution is, to do an inverse filtering to reconstruct the image as discussed earlier. This consists of taking the Fourier transform of (2) so that the convolution operation will become a simple multiplication operation and the deconvolution is just a division operation.

$$\hat{F}\left(k(s,t,s_0,t_0)\right)\hat{F}\left(u(s,t)\right) = \hat{F}\left(z(s_0,t_0)\right) \qquad (3)$$

here $\hat{F}(.)$ denotes 2-D Fourier transform. So the inverse problem can be stated as:

$$u(s,t) = F^{-1}\left(\frac{\hat{F}\left(z(s_0,t_0)\right)}{\hat{F}\left(k(s,t,s_0,t_0)\right)}\right) \quad (4)$$

where $F^{-1}(.)$ is the inverse Fourier transform. In many cases this will not work well. The reason being the values of $\hat{F}\left(k(s,t,s_0,t_0)\right)$ will be negligibly small for high frequency components, which make the denominator a considerably small value that will result in blowing up of the quantity on the RHS of (4). This makes discrepancies in the reconstructed image.

There have been considerable improvements on these types of solutions to the inverse problem, but in general these methods seem to be inadequate to give a proper reconstruction. The deconvolution problem is not well posed in the sense of Hadamard [5], so the solution will not be trivial. This observation made researchers to think in a different perspective for solving these kinds of ill-posed problem, which lead the way to utilize the regularization methods to solve problems of this kind. In this paper we employ a regularization method based on [6].

This paper is organized in six Sections. Section II explains the background of watermarking, Section III explains about the proposed method and the mathematical background for the method. Section IV explains about the numerical implementation scheme employed for the solution of Partial Differential Equations (PDE). Section V highlights on the results and the analysis of the results. Section VI concludes the work.

II. The Background of Watermarking

Digital watermarking technique mainly deals with embedding the digital watermark in digital images. The water mark is added into the images before the images are send to the specified destination [1]. There are a number of ways in which the embedding can be done [2]. The common method followed in digital watermarking technology is spatial domain watermarking [2]. The proposed method exploits this technique. The spatial domain watermarking can be done either locally or globally [7], [8]. Here we are using a global watermarking technique so that the watermark cannot be removed easily. In this proposed method we are assuming the watermark as kernel of a linear operator or impulse response of a Linear Time Invariant system (LTI), (the operation is convolution) and the watermark extraction process as a deconvolution technique, so that the recovered image is similar to the original image. The result is compared visually and quantitatively. The extracted and the original images are compared and the difference image is obtained, as well as the extracted kernel is compared to the embedded one and the error of extraction is within the tolerance limit. The Section V explains the results in detail.

III. Proposed Model

Here we propose a method which uses a Total Variation (TV) deconvolution technique to extract the watermark embedded in the image. The watermark embedding procedure can be mathematically defined as a convolution process as in (2). The blind deconvolution proposed by T.F Chan et. al. [6] is used in our work to reconstruct the image. The TV deconvolution can be mathematically formulated as minimizing the objective function:

$$\min_{u,k} f(u,k) = \|u*k - z\|^2_{L^2(\Omega)} + \alpha_1 \int \|\nabla u\|_{TV}\, dx + \alpha_2 \int \|\nabla k\|_{TV}\, dx \quad (5)$$

with the constraints

$$u,k \geq 0,\ \int k(x,y)dxdy = 1,\ k(x,y) = k(-x,-y) \quad (6)$$

where u is the actual image k is the convolution kernel (watermark) α_1 and α_2 are the parameters which controls the good fit and regularity of the solutions u and k. Further we assume that the convolution kernel has a Gaussian distribution with mean 0 and variance σ^2 and this will result is an out of focus blur in the watermarked image and in many cases this will not be detectable by the intruder and for the intruder the image would just look like a blurred one and the image can only be deblurred by the intended person who has some information regarding the original image or the mask used for creating the blurred image. But generally this method is proposed for the case when there is no information regarding the kernel used for blurring the image. The only handle for the extraction process is, the information that the embedding is a convolution process and the embedded data is of Gaussian distribution of mean zero and variance σ^2. So here we mainly discuss the aspects of *steganalysis* in which the data needs to be recovered without much information regarding the watermark.

IV. Numerical Implementation

From (5) we can obtain an image u and convolution kernel k which minimizes the objective function (5) subject the constraints given in (6). In order to solve (5) which is in dual variable, we have to devise a method to minimize the function for both the variables and hence we can write the Euler-Lagrange equations for the function in (5) as:

$$\frac{\partial f}{\partial k} = u(-x,-y)*(u*k - z) - \alpha_2 \nabla.\left(\frac{\nabla k}{|\nabla k|}\right) = 0\,; \quad x \in \Omega \quad (7)$$

and

$$\frac{\partial f}{\partial u} = k(-x,-y)*(k*u-z)$$
$$-\alpha_1 \nabla.\left(\frac{\nabla u}{|\nabla u|}\right) = 0 \qquad x \in \Omega \tag{8}$$

The equations (8) and (9) are solved simultaneously using Alternating Minimization method (AM) [6]. Homogeneous Neumann boundary conditions $\left(\frac{\partial u}{\partial \vec{n}} = 0\right)$ are assumed for the PDE's defined throughout the paper (where $\vec{n}$ is the unit outward normal). We have used the explicit Euler scheme for solving the PDE's numerically. The details can be found in [6]. The values of α_1 and α_2 can directly affects the minimization process. The parameter α_2 controls the spread of the kernel. That is, if α_2 is large then the spread of the kernel is big whereas α_1 is directly proportional to the Signal to Noise Ratio (SNR) of the image. The differential equations in (7) and (8) are discretized using the central difference scheme. The term $\nabla.\left(\frac{\nabla u}{|\nabla u|}\right)$ is discretized using the central difference scheme as:

$$\frac{I_{xx}I_y^2 - 2I_{x,y}I_xI_y + I_{yy}I_x^2}{(I_x^2+I_y^2)^{\frac{3}{2}}} \tag{9}$$

where

$I_{xx} = \frac{I_{i+1,j} - 2I_{i,j} + I_{i-1,j}}{h^2}$, $I_{yy} = \frac{I_{i,j+1} - 2I_{i,j} + I_{i,j-1}}{h^2}$,

$I_x = \frac{I_{i+1,j} - I_{i-i,j}}{2h}$, $I_y = \frac{I_{i,j+1} - I_{i,j-1}}{2h}$ and

$I_{xy} = \frac{I_{i+1,j+1} - I_{i+1,j-1} - I_{i-1,j+1} + I_{i-1,j-1}}{4h^2}$. Here we assume the scale space parameter $h = 1$.

Experimentally it is found that the spread of Point Spread Function (PSF) is directly proportional to α_2 so when α_2 is around 10^{-6} the area of support of PSF is very less and the recovered image is sharper keeping α_1 fixed as $3X10^{-6}$. The iterations are carried out till convergence and the pre-conditioner in [9] is used to get a better convergence and the iterations are carried out until the residual becomes 0.15. The conjugate gradient method is adopted to solve this linear system.

V. Results and Discussions

Testing was done on different images of size 512X512 the Gaussian random array of size25X25 was used as a kernel and the convolution procedure was carried out to embed the data into the image. The image will get blurred a little when the data gets inserted and the data inserted will remain un-identified to the naked eye. The Fig 1 shows the result images, Fig 1(a) is the original image, Fig 1 (b) is the Gaussian kernel used as the watermark the kernel size is 25X25 with zero mean and variance σ^2 = 5. Figure 1(c) shows the watermarked image and it can be easily seen that the added watermark will not be distinguishable to the naked eyes. Fig 1 (d) shows the extracted watermark and Fig. 1(e) is the extracted image. Fig 1(f) is the difference image and it can be seen that some of the edge information is lost during the watermarking process so the deblurring operation will tends to lose some details which are affordable in the watermarking scenario. The *steganalysis* or the watermark extraction process is literarily a deblurring process in which the watermark embedded is extracted with a desired accuracy and the image extracted is comparable to the original image. The method holds good even if the images are corrupted by the Gaussian noise. In other words the method can extract the watermarks even if the images are noisy.

VI. Conclusion

In this paper we have proposed a novel method to secure digital data through digital watermarking. The watermarking is assumed to be a linear operation with a bounded linear operator and the watermark extraction process is a de-convolution process. Since the de-convolution process is highly ill-posed in nature we have employed a regularization method to extract the watermark from the image. The method was implemented and tested for variety of images and the results provided endorses on the efficiency of the method to carryout *steganalysis* with a better accuracy.

Acknowledgment

The authors wish to thank the National Institute of Technology for providing financial support through Seed money grant-2009-10. The authors also wish to record the sincere appreciation extended by the research community of the Institute.

References

[1] Niel F Jhonson, An intruduction to watermark recovery from images, Proceedings of the SANS Intrusion Detection and Response Conference, (1999).

[2] R. Anderson and F. Petitcolas. On the limits of stehanography. IEEE Journal of Selected Areas in Communications, No. 16, Vol.4, pp. 474.481, May (1998).

[3] A.I. Hashad, A.S. Madani and A.E.M.A. Wahdan, A robust steganography technique using discrete cosine transform insertion, in: Proceedings of IEEE/ITI 3rd International Conference on Information

and ommunications Technology, Enabling Technologies for the New Knowledge Society, pp. 255-264, Dec. (2005).

[4] N. Meghanathan and Lopamudra Nayak, Steganalysis algorithms for detecting the information in image audio and video cover media, International journal of network security and its applications, No. 1, Vol. 2, Jan (2010).

[5] Hardmard J, "Lecturer on Cauchy's, problem in linear partial differential equations" , Dover publication, (1953).

[6] Tony F. Chan, Chiu-Kwong Wong: Total variation blind deconvolution. IEEE Transactions on Image Processing No. 7, Vol. 3, pp. 370-375, (1998).

[7] Christian Cachin. An information-theoretic model for steganography, Lecture Notes in Computer Science, pp. 306-318, (1998).

[8] K. Bailey and K. Curran, An evaluation of image based steganography methods, Multimedia Tools andApplications, 30, Vol. 1, pp. 55-88,(2006).

[9] Raymond H. Chan, Tony F. Chan, Chiu-Kwong Wong: Cosine transform based preconditioners for total variation deblurring. IEEE Transactions on Image Processing Vol. 8(10), pp. 1472-1478, (1999).

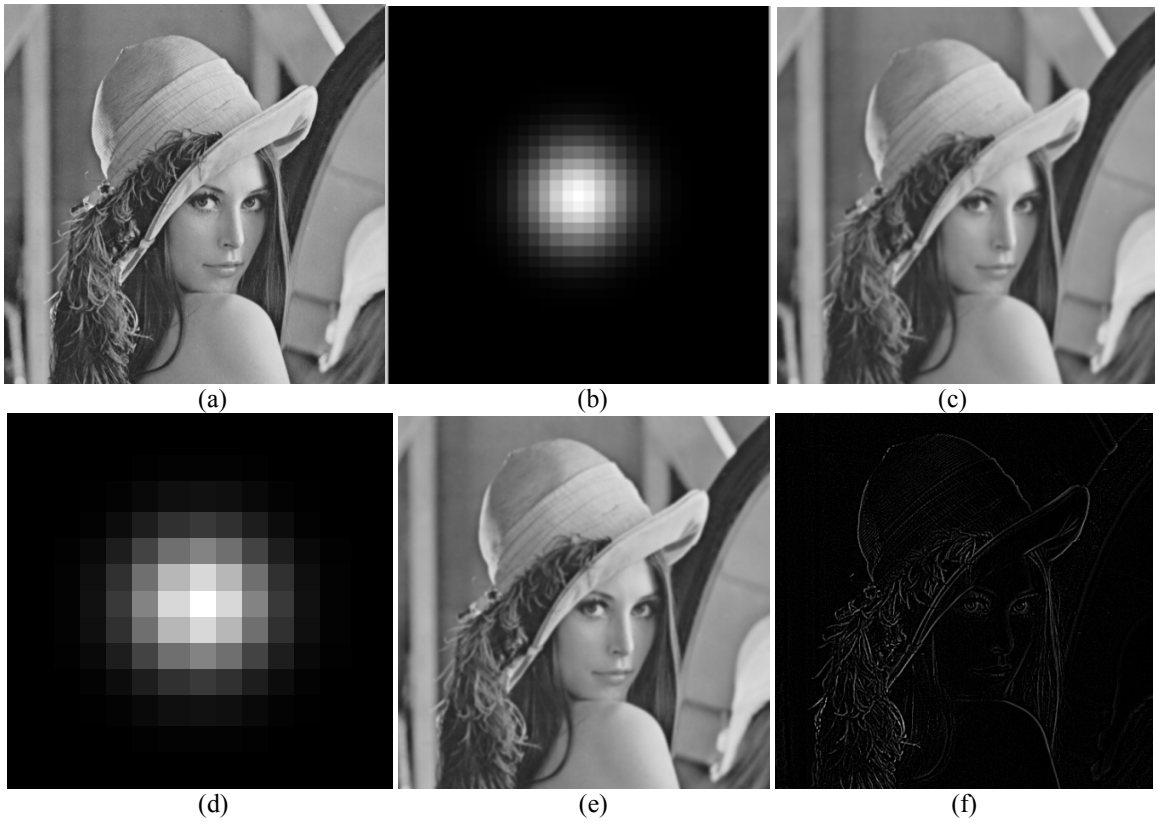

Figure 1. (a) The original image (b) the watermark (Gaussian Kernal with size 25X25.) (c) the watermarked image (d) the extratced watermark using the proposed method (e) the extracted image (f) the difference image.