2014

Steganography Attack Based on Discrete Spring Transform and Image Geometrization

Qilin Qi
University of Nebraska-Lincoln

Aaron Sharp
University of Nebraska-Lincoln

Yaoqing Lamar Yang
University of Nebraska-Lincoln, yyang3@unl.edu

Dongming Peng
University of Nebraska-Lincoln

Hamid Sharif
University of Nebraska-Lincoln, hsharif@unl.edu

Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International
Year: 2014
Pages: 554 - 558, DOI: 10.1109/IWCMC.2014.6906416

Steganography Attack Based on Discrete Spring Transform and Image Geometrization

Qilin Qi, Aaron Sharp, Yaoqing Yang, Dongming Peng, Hamid Sharif
{atsharp, qqi}@unomaha.edu, {yyang2, dpeng, hsharif}@unl.edu
Department of Computer and Electronics Engineering
University of Nebraska-Lincoln

***Abstract*—In order to prevent harmful secret information sharing by steganography, a new active-warden countermeasure approach against steganography is proposed in this paper. Differently from the other countermeasure approaches presented in current literature which need to have some knowledge of the steganographic algorithms to be attacked, our proposed method is a generic method which is independent of any particular steganography methods being utilized. In other words, our approach blindly attacks the steganography without any prior knowledge of the used algorithms or their existence. In general, the hidden information is embedded in a carrier by adjusting the coefficients of the audio or image. In our method, by exploiting the large margin between the numerical value and visual perception of the images, large amounts of visually non-detectable distortions are incurred in the image. As a result, the hidden message is destroyed by this method while the perceptual quality of the image is maintained. Inspired by the print-scan process in which most of the steganographic methods cannot survive, a transform called Discrete Spring Transform (DST) is proposed in this paper as the foundation of our attack algorithm. An image geometrization method is also developed to reconstruct the image in this paper. The simulation results have demonstrated that the PSNR of the attacked image is above 30dB with a high perceptive quality while the BER of the hidden steganographic message in the attacked image is above 0.5.**

Keywords—steganography; multimedia security; print-scan process; human visual system

I. INTRODUCTION

Security is a rising problem in wireless communication and networks. Steganography is a potentially strong method to transmit secret information especially by hiding messages in the multimedia data through networks [1]. Though steganography provides a way to transmit hidden information, however, the malicious use of steganography draws another stringent security problem. Evidences have shown that terrorists use steganography to broadcast operation instruction and training manuals on the Internet [2]. Consequently, preventing malicious information transmission over Internet is a critical issue for public security. The prominent difficulty is that the steganography will not draw any attention. In other words, it is difficult to perceptually identify whether an image/video/audio is hidden with some information or not. Furthermore, it is even hard or impossible to detect which steganography method is used since there are hundreds of existing steganography methods.

Steganalysis is a kind of countermeasures to steganography[3]. Steganalysis tries to detect the 1.69steganography and detect what information is embedded. Many effective steganalysis methods are proposed in the past decade [4-8]. Steganalysis may be very useful for analyzing a highly suspicious image where some hidden information is believed to be embedded with a known steganography method. However, most of the steganalysis is only valid for a certain kind of steganography method because the mechanism of the steganalysis is based on the specific statistic features of steganography algorithms. As discussed above, there may be millions of images transmitted over Internet some of which may contain hidden messages by hundreds of different steganography methods. Therefore, it is extremely difficult to detect which kind of steganography method is used for a certain image. Then it is also difficult to locate a steganalysis method targeting that steganography method. In fact, over Internet, most of the information is secure where no hidden information is embedded, so huge resources will be wasted if all the images over Internet are tested by many different kinds of steganalysis methods. In all, steganalysis as a passive and non-generic countermeasure to steganography is not proper for the Internet environment.

In [11], an active steganalysis method is proposed against QIM steganography. It is an active method which is capable of detecting whether QIM steganography exists in the cover media or not using a measure of entropy which is called approximate entropy. It has advantages over passive steganalysis which is not able to detect the existence of such steganography. However the proposed method can only work against QIM steganography. In [12], a learning based steganalysis method is proposed. Instead of using the common SVM scheme, here an ensemble classifier is utilized in [12] to achieve a lower complexity learning process. Nevertheless, the learning process requires a large amount of test data and only works on known steganography methods. In [13] and [14], active warden approach is discussed to eliminate steganography in the Internet traffic. The potential advantages for active warden methods are investigated. In [15], [16] and [17] are some specific active warden based attacking approaches against video, audio and image steganography. All those methods attempt to remove the steganography in the cover media while preserving the perceptual quality of the cover media, but they all require some prior knowledge to some extent.

A generic active countermeasure method is desired to be used in many real world scenarios where the prior knowledge of the steganographic algorithms may not be available at all. In this paper, we assume all the images transmitted through a server on the Internet are applied by our proposed method. The visual perceptual quality of the

output images should be at an acceptable level which means that no significant visual distortion can be found in the output images. In the same time, the hidden information embedded in the compromised images are securely removed. It doesn't necessarily mean the hidden message is totally destroyed. It will be found in the next section of this paper that rather than destroying the hidden information, the position of the hidden information is disordered so that the receiver is not able to recover the hidden information. Though the exact hidden information is not able to be discovered as in the steganalysis method, the malicious information spreading will be prevented and the network can be kept at a high security level.

In order to find such a generic method, the steganography attack scheme should not be based on any specific steganographic mechanisms. In general, the steganography methods take advantage of the human visual system to hide information. Due to the limitation of the human visual system, by properly designing, the slight changes of the image's numerical values either in spatial domain or transform domain will not draw significant perceptible distortions [7]. This flexibility of the image values provides a capacity to carry hidden information. The countermeasure also can take advantage of the human visual system. In our proposed method, we try to maximize the distinction between the human visual perception and the numerical value of the image to make the maximum numerical change to the image while keeping the image visually unchanged. The significant numerical change of the image will make the hidden message difficult to be recovered. It should be noted that since the idea behind the proposed method is to change the numerical value of the image, the traditional Mean-Square Error (MSE) measurement is not fair to evaluate the image quality of the output image applied by the proposed method. In section V, a subjective image evaluation method is adopted to validate the results.

The print-scan is a physical process which generally meets the properties of the countermeasure method as mentioned above. The state-of-art printer and scanner are capable of presenting and recovering a digital image in a very high visual quality. However, the numerical values of the rescanned image may be largely different from the original one. Meanwhile, it is also demonstrated that most of the steganography method is not robust to the print-scan process [15]. Even though there are a few data hiding algorithms which can survive the print-scan process, the data-hiding bandwidth or capacity is severely constrained. Inspired by the print-scan process, a countermeasure method called Discrete Spring Transform (DST) is elaborated in section II and section III. A preliminary consideration of DST has been initially reported in our previous work [15]. Such DST can be tweaked as a countermeasure to steganography which is even stronger than physical print and scan. In fact, a major distortion introduced by print-scan process is geometric distortion to which human visual system is not sensitive. It also looks like a spring which are stretched a little bit. In every point of the spring, it is stretched at a different rate. But the entire spring is stretched smoothly where the specific stretching difference in a single point is non-noticeable. To make this geometric deviation smarter, an Analog Location Transform (ALT) is proposed to make the deviation rate inversely proportional to the visual significance of the image area. A geometrized image DST method is proved to be effective by the simulation results.

The rest of the paper is organized as follows. The discrete spring transform is presented in section II. In section III, geometrized image reconstruction method is described. The simulation results are shown in section IV. Finally, we draw conclusion in section V.

II. DISCRETE SPRING TRANSFORM

A. Analog Location Transform

A major reason why most of the steganography methods fail to survive from the print-scan process is the geometrical location deviation distortion. The mechanical structure of the printer and the aligning error during the print-scan process will easily make the original pixels deviate from their original locations. This deviation will not remove the hidden information embedded in the pixels but it will let the receiver unable to locate and synchronously decode the hidden information. First of all, the random and variant deviation distance is very difficult to be found. Meanwhile, since the deviation in each pixel is relatively small and random, this deviation distortion cannot make a general impact on the perceptual quality of the image. Human visual system is not sensitive to this geometrical distortion. This physical pixel location deviation can be virtually modeled as an analog location transform (ALT). A single dimensional signal Discrete Spring Transform can be illustrated in Fig.1 where the dash line denotes the original locations of the signal.

For a M×N image $A(m,n)$, the ALT-based Discrete Spring Transform is presented as

$$A(m,n) \to A'(m',n')$$

where

$$\begin{pmatrix} m' \\ n' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta_1(m,n) \\ 0 & 1 & \Delta_2(m,n) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ n \\ 1 \end{pmatrix}$$

$$(m,n \in \mathbb{Z}\ \ m',n' \in \mathbb{R})$$

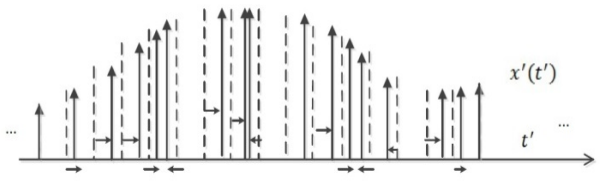

Fig.1 1-D Discrete Spring Transform

This transform is a spatial-variant transform which differs from a conventional affine transform or geometric transform. Please note the variant of the deviation distance $\Delta_1(m,n)$ and $\Delta_2(m,n)$ is a key reason why the hidden message in the attacked image is unable to be recovered. In order that the deviation does not mess up with the relative postion of the pixels, the transform functions are confined as $|\Delta_1,\ \Delta_2| \le 0$ and $\forall\, m = 0$ or $n = 0, 0 \le \Delta_1(m,n), \Delta_2(m,n)$, $\forall\, m = \mathrm{M}-1$ or $\mathrm{n} = \mathrm{N}-1,\ \Delta_1(m,n), \Delta_2(m,n) \le 0$. This condition makes sure that the location of each pixel is deviated only in relation to its nearby area. The probability density function can be expressed as

$$\mathrm{P}\big(\Delta_i(m,n)\big) = \begin{cases} \dfrac{1}{2\gamma_\mathrm{i}}, & -\gamma_\mathrm{i} < \Delta_i(m,n) < \gamma_\mathrm{i} \\ 0, & \text{otherwise} \end{cases}$$

where $\mathrm{i} = 1{,}2$ and $0 < \gamma_\mathrm{i} < 0.5$ is the variable range of the deviation distance. This maximum deviation distance provides a reference threshold for the following method where the deviation distance is associated with the properties of the human visual system. These thresholds will then be used for normalizing the deviation distance functions.

B. Curve Length Method

In order to remove the hidden information from the image, the deviation distance is expected to be as large as possible to result in numerical distortions. On the other hand, larger distortions will obviously deteriorate the image quality. Generally, the distortion is preferred to be involved in the image areas where the human visual system is not sensitive. It is also reasonable that those areas are more suspicious to have hidden information because the change is less noticeable. For example, human visual system is more likely to notice the contrast of the pixels than the pixel intensity itself [10]. Consequently, the area with edges of objects within the image has less tolerance for the distortions than the plain area. Then larger deviation distance needs to be introduced in the plain area. In this paper, we propose a so-called "curve length" method to model this distinction in the image between the high contrast areas and the plain areas, and increase or suppress the deviation distance accordingly. The curve length method is capable of progressively and smoothly manipulating the transition from the plain area to the edge area.

To calculate the curve length map for an image, the image is interpolated at first. The 2-D interpolated signal can be expressed as

$$(m,n) = A(m,n) * W_L(m,n)$$

where $\mathrm{W_L}$ is the interpolation window kernel which is a 2-D extension of Lanczos window kernel. The curve length between $A(k,n)$ and $A(k+1,n)$ horizontally and between $A(m,k)$ and $A(m,k+1)$verticaly can be presented as

$$l_1(\mathrm{k},n) = \int_{\mathrm{x=k}}^{\mathrm{x=k+1}} \left(\frac{\mathrm{d}\hat{A}}{\mathrm{dx}}(\mathrm{x},n)^2 + 1 \right) \mathrm{dx}$$

$$l_2(m,\mathrm{k}) = \int_{\mathrm{y=k}}^{\mathrm{y=k+1}} \left(\frac{\mathrm{d}\hat{A}}{\mathrm{dy}}(m,\mathrm{y})^2 + 1 \right) \mathrm{dy}$$

In other words, the expression above shows that we can manipulate the deviation of the pixel positions in the 2-D plane in such a way that the deviation of position is related to the pixel values.

After the position deviation process, each pixel is moved to a new location where the coordinates may not be integral due to the nature of analog location transform. A novel pixel geometrization method is proposed in next section to reconstruct the image pixels by virtually implementing the scanning process.

III. GEOMETRIZED IMAGE RECONSTRUCTION

Assume a scanner will scan the image back into digital form based on the image which is virtually printed by the ALT method. The virtual ink dot position is overlapped due to the pixel position deviation and ink spreading. The ink volume for each pixel is assumed to be proportional to the pixel's value. As a result, a larger ink volume dot will be spread in a larger circle area which can be denoted as

$$A'(\mathrm{x}',\mathrm{y}') = \pi \mathrm{r}(\mathrm{x},\mathrm{y})^2$$

where *r* denotes the radius of the ink dot circle. Considering this ink spreading phenomenon and the deviated ink dot center point, it is easy to express and manipulate the distortion to the virtually printed image. For the scanning process, a square-shaped scanning window with side length *d* is defined. The scanning window slides in the 2-D plane vertically and horizontally and generates the digitized pixel values at discrete locations. The ink volume captured in the scanning window is quantified to reconstruct the image pixels. Each ink dot will have impact on a group of neighbor pixels. Though the overlap between the scanning window and ink dot can be modeled in many different patterns, it can be formulated with a concise expression. Fig.2 illustrates some examples of the scanned image pixels corresponding to the ink dots. The arrows indicate the deviation directions of ink dots and the ink dot size is propotional to the pixel value.

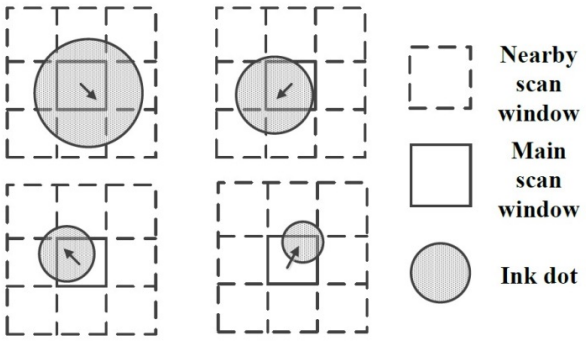

Fig.2 Examples of Geometrized Image Representation

As shown in Fig.2, one ink dot may be spread into eight nearby scanning windows. Therefore nine pixel values are affected by this dot after the virtual scanning transformation. The ink volume spread by $A(i,j)$ to the neighbor pixels can be expressed as

$$S^{(i,j)}(i \pm 1, j \pm 1) = P^{r(x,y)}\left(\frac{d}{2} \mp \Delta_x(i,j), \frac{d}{2} \mp \Delta_y(i,j)\right)$$

$$S^{(i,j)}(i \pm 1, j) = T\left(\frac{d}{2} - \Delta_x(i,j)\right) - S^{(i,j)}(i \pm 1, j-1) - S^{(i,j)}(i \pm 1, j+1)$$

$$S^{(i,j)}(i, j \pm 1) = T\left(\frac{d}{2} - \Delta_y(i,j)\right) - S^{(i,j)}(i-1, j \pm 1) - S^{(x,y)}(i+1, j \pm 1)$$

$$S^{(i,j)}(i,j) = \pi r^2(i,j) - \sum_{p=-1}^{1}\sum_{q=-1}^{1} S^{(i,j)}(i+p, j+q)$$

where

$$T(x) = \left[r^2 \arccos\left(\frac{x}{r}\right) - x\sqrt{r^2 - x^2}, 0\right]^+$$

expresses the area in a circle divided by a chord and

$$P(x,y) = \left[\frac{1}{2}\big(S(x) + S(y)\big) + xy - \frac{\pi r^2}{4}, 0\right]^+$$

presents the area in the middle of the two intersecting chords. Then the reconstructed image or the virtually scanned image can be expressed as

$$A_r(i,j) = \sum_{q=j-1}^{j+1}\sum_{p=i-1}^{i+1} S^{(p,q)}(i,j)$$

IV. Simulation Results

The spread spectrum steganography is used in this section to evaluate the proposed method. In experiments, 32 bit-information is hidden in the 256 by 256 images. Three different categories images are tested in this section which are textures, portraits and animals. Thirty pictures in each category are tested by the proposed method.

A typical result of the proposed attacking method based on the DST and virtual print-scan scheme is illustrated in Fig. 3 with a comparison against the original input image. It demonstrates that the artifact of the proposed attack is perceptually unnoticeable, even though the hidden message has been successfully removed from the image. In the attacked picture, the BER (Bit Error Rate) of the hidden information is 0.5 which means the receiver is not able to decode the hidden information with a performance better than the random guess.

In our following experiments, the robust spread spectrum based steganography which is one of the current main stream steganographic methods is targeted for attack. The performance data including BER of the decoded stego-data and PSNR (Peak Signal to Noise Ratio) of the processed images have been collected to evaluate and validate the proposed attacking method. In our experiments, a 32-bit secret information is hidden in the 256 by 256 images. Three different categories of image are tested in this section which include the images featured with textures, portraits and animals. The simulation results shown in Fig. 4 are averaged values. In Fig. 4, the proposed attacking method is evaluated in two aspects including the strength of the attack and the quality of the processed images. In order to meet the requirement of removing the steganography, the BER of the hidden message needs to be close to 0.5. In Fig. 4 (a), the side length of the attacking window which is defined as the scanning window is changed from 4 to 9. It is shown that the attack is not effective when the side length of the window is greater than 6. This is because the scanning window allows most of the geometrized pixels are gathered into the original scanning window. In other words, the virtual scan process does not take effect, and the image data are basically the same as those in the original image. Therefore less distortions are involved between the neighboring pixels, resulting in that the stego-data are untouched. When the window side length is less than 5, however, the BERs become much greater. On the other hand, Fig. 4(b) shows that the PSNR of each category of the pictures is above 30dB when the window side is 4 to 5, meaning that the processed images have been protected to maintain a satisfactory visual quality.

To sum up, the proposed steganography attacking method can be very effective in terms of removing the hidden message while protecting the hosting image quality. The attacking window size is a critical perimeter in designing the attack algorithm. Our experiments have demonstrated that the algorithm has the best performance when the side length of the window is 4 to 5 when the images are of size 256 by 256.

V. Conclusion

In this paper, we have investigated on how to attack the steganography in digital images without degrading perceptual qualities of the host images. We have proposed a novel approach called Discrete Spring Transform and applied this approach into a virtual print and scan process, in which the images were virtually converted into different domains. Our experimental results have demonstrated that the proposed method is very effective in removing the stego-data, with its BER as high as 0.5 while the qualities of the host images are satisfactorily maintained with PSNR mostly higher than 30dB.

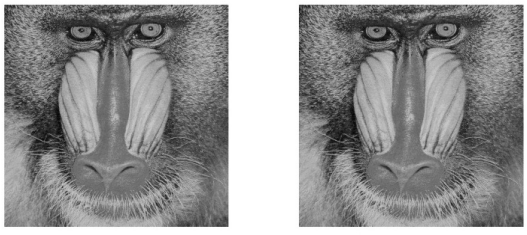

Fig. 3 Comparison between Original and Attacked Image

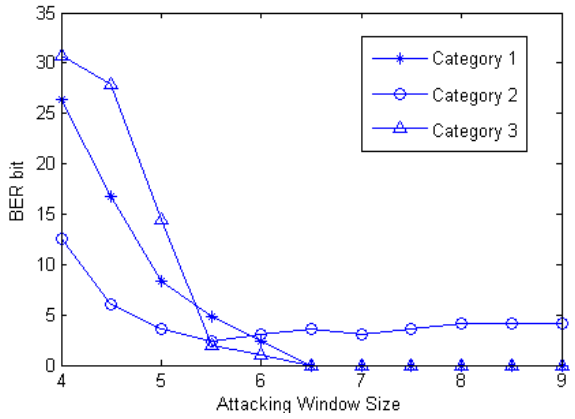

Fig. 4(a) BER of the Attacking Results for Different Picture Categories

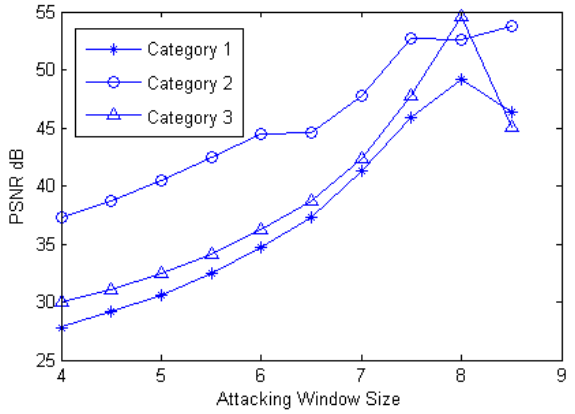

Fig. 4(b) PSNR of the Attacking Results for Different Picture Categories

ACKNOWLEDGMENT

This paper was funded in part by the Nebraska Research Initiative on Multimedia Rendering.

REFERENCES

[1] N. Provos and P. Honeyman, "Hide and seek: an introduction to steganography," Security & Privacy, IEEE, vol. 1, pp. 32-44, 2003.

[2] Interagency Working Group on Cyber Security and Information Assurance, "Federal Plan For Cyber Security And Information Assurance Research And Development," 2006.

[3] R. Chandramouli, M. Kharrazi, and N. Memon, "Image Steganography and Steganalysis: Concepts and Practice." vol. 2939, T. Kalker, I. Cox, and Y. Ro, Eds., ed: Springer Berlin / Heidelberg, 2004, pp. 204-211.

[4] L. Bin, F. Yanmei, and H. Jiwu, "Steganalysis of Multiple-Base Notational System Steganography," Signal Processing Letters, IEEE, vol. 15, pp. 493-496, 2008.

[5] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB Steganography via Sample Pair Analysis Information Hiding." vol. 2578, F. Petitcolas, Ed., ed: Springer Berlin / Heidelberg, 2003, pp. 355-372.

[6] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," presented at the Proceedings of the 2001 workshop on Multimedia and security: new challenges, Ottawa, Ontario, Canada, 2001.

[7] H. Malik, K. P. Subbalakshmi, and R. Chandramouli, "Steganalysis of QIM-based data hiding using kernel density estimation," presented at the Proceedings of the 9th workshop on Multimedia & security, Dallas, Texas, USA, 2007.

[8] L. Zhuo, L. Kuijun, Z. Xianting, and P. Xuezeng, "Feature-Based Steganalysis for JPEG Images," in Digital Image Processing, 2009 International Conference on, 2009, pp. 76-80.

[9] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," Signal Processing, vol. 70, pp. 177-200, 1998.

[10] D. C. Hood and M. A. Finkelstein, "Sensitivity to light," in Handbook of perception and human performance. vol. 1, ed New York: Wiley-Interscience, 1986, pp. 5-1 to 5-66.

[11] Malik, H.; Subbalakshmi, K. P.; Chandramouli, R., "Nonparametric Steganalysis of QIM Steganography Using Approximate Entropy," Information Forensics and Security, IEEE Transactions on , vol.7, no.2, pp.418,431, April 2012

[12] Kodovsky, J.; Fridrich, J.; Holub, V., "Ensemble Classifiers for Steganalysis of Digital Media," Information Forensics and Security, IEEE Transactions on , vol.7, no.2, pp.432,444, April 2012

[13] Zawawi, M.N.; Mahmod, R.; Udzir, N.; Ahmad, F.; Desa, J.M., "Active warden as the main hindrance for steganography information retrieval," Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on, vol., no., pp.277,280, 13-15 March 2012

[14] Gina Fisk, Mike Fisk, Christos Papadopoulos, and Joshua Neil. 2002. Eliminating Steganography in Internet Traffic with Active Wardens. In Revised Papers from the 5th International Workshop on Information Hiding (IH '02), Fabien A. P. Petitcolas (Ed.). Springer-Verlag, London, UK, UK, 18-35.

[15] Sharp, A.; Qilin Qi; Yaoqing Yang; Dongming Peng; Sharif, H., "A novel active warden steganographic attack for next-generation steganography," Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International , vol., no., pp.1138,1143, 1-5 July 2013

[16] Qi, Qilin; Sharp, Aaron; Peng, Dongming; Yang, Yaoqing; Sharif, Hamid, "An active audio steganography attacking method using discrete spring transform," Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on , vol., no., pp.3456,3460, 8-11 Sept. 2013

[17] Rezaei, F.; Hempel, M.; Shrestha, P.L.; Tao Ma; Dongming Peng; Sharif, H., "A quality-preserving hidden information removal approach for digital images," Communications (ICC), 2012 IEEE International Conference on , vol., no., pp.1021,1025, 10-15 June 2012