# Blind decoding of image steganography using entropy model

C.R. Kim, S.H. Lee✉, J.H. Lee and J.-I. Park

Image steganography hides secret information in the cover images so naturally that the existence of hidden data in the stego-image is not recognisable. This Letter proposes a new approach to blind decoding of image steganography using the local entropy distributions of decoded images. The local entropy distributions of incorrectly decoded images are different from those of normal ones because of the abnormal image structures in the erroneously decoded images. This blind decoding in the image steganography is very useful to extract hidden image information because there are enormous least significant bit (LSB)-based steganography methods, and it is very hard to find the methods by observing manipulated LSBs.

*Introduction:* Steganography is a technique to hide secret information in the multimedia such as image, video, audio signals etc. Especially, image steganography hides much information in the cover images so naturally that the other users do not recognise the existence of hidden data when looking at the stego-images [1–3].

There are two categories in image steganography, frequency-domain approaches, and spatial methods, respectively. The frequency-domain approaches usually handle with discrete cosine transform (DCT) coefficients in JPEG encoding. Jsteg, F3, F4, and F5 algorithms are the steganography encoding algorithms to manipulate DCT coefficients in hiding data [4, 5]. The steganography methods in the frequency domain show usually low capacity to hide data and high distortion of stego-images. On the other hand, the spatial methods have larger data capacity and better visual quality of stego-images. The small change of each pixel is not noticeable by the human inspection. Specifically, image steganography using least significant bit (LSB) manipulation is to change the LSBs of pixels to embed the information [6–10]. The bits of a pixel in the information image are divided into 2 bits and embedded into the LSBs of cover image pixels. Thus, the visual difference between pure cover image and stego-image is not noticeable.

This Letter focuses on spatial image steganography and its blind discrimination. The proposed approach of steganography discrimination is different from the usual steganalysis approaches where they determine an image to be stego-image or not by inspecting the LSBs and other cues [11, 12]. The proposed method is to discriminate the correctness of stegano-decoding results. Consequently, the proposed approach is to find the correct encoding method and the information image simultaneously. Even though we recognise the suspicious image is encoded by a steganography method, it is another work to find the exact encoding method and hidden information in the stego-image. This Letter proposes a blind decoding approach to find the hidden image information by observing local entropy distributions in the extracted images.

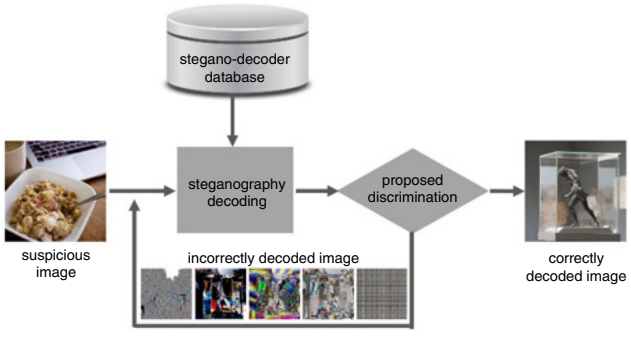


**Fig. 1** *Proposed approach of blind steganography discrimination. Correct decoding method and hidden image information are found by determining decoded resulting image to be normal image*

*Proposed approach for blind steganography decoding:* The proposed approach for blind steganographic decoding is depicted in Fig. 1. Assume that the libraries of encoding/decoding methods are given. A suspicious image is first decoded by a decoding method in the libraries and is tested by the proposed discrimination method. The discrimination method analyses the local entropy distribution of the decoded image. Since incorrectly decoded images do not show the normal image structures, the local entropy distributions of wrongly decoded images are different from those of normal images. If the decoded image is determined to be incorrect by the entropy characteristics, the next decoding method is applied to the suspicious image and tests the correctness. This process is repeated until the decoding result is determined to be correct. Since there are a huge number of steganographic encoding methods in the LSB manipulations [8], it is almost impossible to check every decoding result by human inspection. The proposed approach discriminates the correct stegano-encoding method and hidden image information by determining the decoded resulting image to be a normal image.

*Entropy models:* This Letter has observed the different characteristics of local entropy between wrongly decoded images and normal images. Fig. 2 shows some incorrectly decoded images from the stego-images using various LSB manipulations. The two LSBs of each pixel in the cover image are changed by 2 bits of a pixel in the information image [7–10]. As shown in Fig. 2, the extracted information images are not normal when the stego-images are incorrectly decoded. When we do not know the exact encoding method, the decoded images are different from the normal images. In the Letter, the visual abnormality is modelled by characteristics of local entropy.

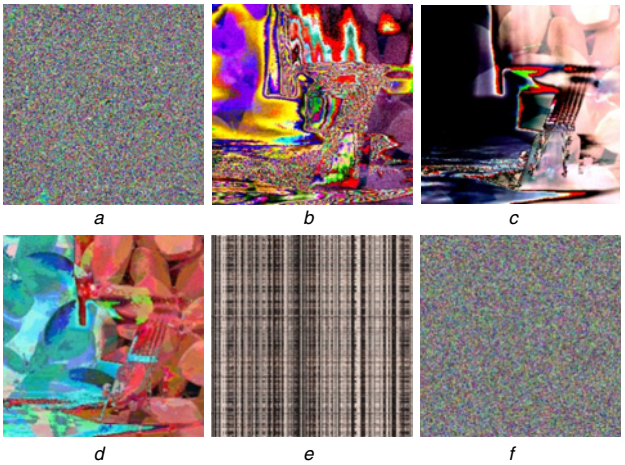


**Fig. 2** *Examples of incorrectly decoded images from stego images without encoding method*

*a* Decoded image from not encoded cover image (Simple)
*b* Disordering pixels of stego-images (DO)
*c* Pixel subtraction and division by 5 (DVI5)
*d* XOR of cover and information images (XOR)
*e* Random permutation of pixels (RP)
*f* Pixel value difference image (PVD)

This Letter has found that the mean and variance of local entropies are a good clue to discriminate the abnormality of decoded images. We observed the difference of entropy characteristics between wrongly decoded images and normal images. Fig. 3 shows the entropy distributions of normal images and incorrectly decoded images for various encoding methods. The entropy is calculated in every local block for Y colour component, and its mean and variance are obtained for an image. As is shown in Fig. 2, the wrongly decoded images are similar to random noise images, thus they have high mean and low variance of entropy. As we can see in Fig. 3, the local entropy distribution is a good feature to determine if the decoded image is a normal image. Note that the distributions of local entropy are different with respect to the encoding methods. Thus, it is desirable that the discrimination rule is separately trained.

This Letter uses the support vector machine (SVM) and Gaussian kernel to classify the normal images and wrongly decoded images. The SVM learned 6D vectors of entropy mean and variance for YCbCr colour components. From the various experiments, the optimal block size is $16 \times 16$ for a test database.

*Experimental results:* This Letter collected 2000 images for obtaining entropy distributions and training discrimination rule. The mean and variance of entropy are calculated with every $16 \times 16$ block for Y, Cb, and Cr components, respectively. Also, the 6D vectors are trained with Gaussian kernel SVM. We experimented with 1000 incorrectly decoded images and

1000 normal images which were randomly selected. This Letter repeated the process of learning and testing 10 times and calculated the average of test results. The results are shown in Table 1. According to the tests, the proposed method determines if the image is correctly decoded with 92.9% accuracy. Some errors shown in Fig. 4 usually occur when the images are very complex like incorrectly decoded images. Since many test images consist of complex structures and randomised patterns, the local entropy characteristics are similar to those of wrongly decoded images. Thus, it is expected that the proposed method works better for the usual images. When the discrimination rule is adjusted to reduce true-negative errors (that is, a normal image is considered as abnormal), the proposed method is more useful for blind decoding from the huge number of steganography methods.

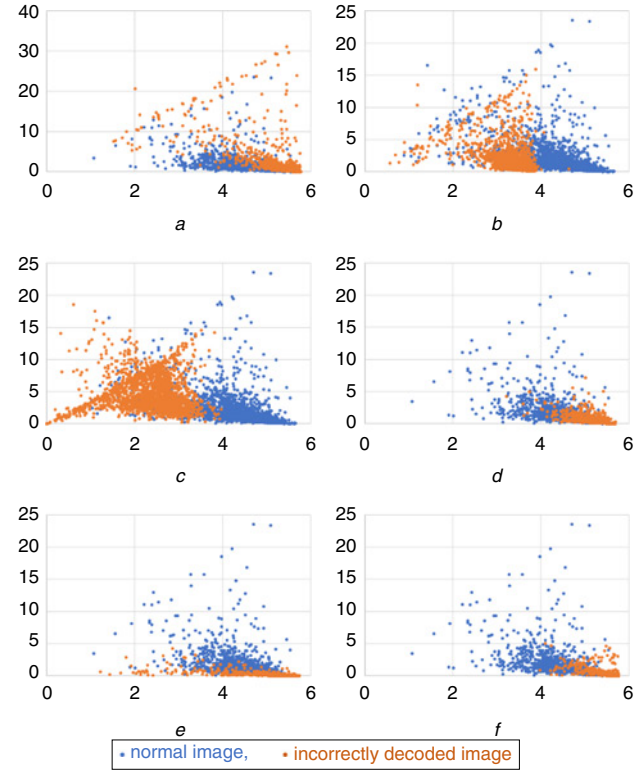


**Fig. 3** *Local entropy distributions of various encoding methods. X-axis is the mean and y-axis is variance of block entropy*

*a* Decoded image from not encoded cover image (Simple)
*b* Disordering pixels of stego-images (DO)
*c* Pixel subtraction and division by 5 (DVI5)
*d* XOR of cover and information images (XOR)
*e* Random permutation of pixels (RP)
*f* Pixel value difference image (PVD)

**Table 1:** SVM discrimination result table of proposed method

| Name | Correct (# of images) | False-positive | True-negative | Correct rate, % |
|---|---|---|---|---|
| Simple | 1848 | 111 | 41 | 92.4 |
| DO | 1865 | 33 | 102 | 93.3 |
| DVI5 | 1895 | 35 | 70 | 94.8 |
| XOR | 1656 | 169 | 175 | 82.8 |
| RP | 1857 | 99 | 44 | 92.9 |
| PVD | 1913 | 72 | 16 | 95.7 |

**Fig. 4** *Some errors in discriminating incorrect decoding. Misidentified images are usually very complex like random images*

*Conclusions:* This Letter has proposed a blind decoding method of image steganography. The local entropy characteristics of normal and wrongly decoded images have been observed and trained using SVM. The proposed method recognises the correct decoding results by an entropy-based model. According to the various experiments, the proposed method detects the correctly extracted images with 92% accuracy. Since there are too many methods for image steganography, it is impossible to check the correctly decoded images for the human supervisors one by one. Thus, the proposed method is very useful for blind decoding of image steganography without any supervisors.

*Acknowledgments:* This work was supported by the research fund of the Signal Intelligence Research Center supervised by the Defense Acquisition Program Administration and the Agency for Defense Development of Korea.


Submitted: *18 November 2017* E-first: *12 April 2018*
doi: 10.1049/el.2017.4276
One or more of the Figures in this Letter are available in colour online.

C.R. Kim, J.H. Lee and J.-I. Park (*Department of Computer and Software, Hanyang University, Seoul, Republic of Korea*)

S.H. Lee (*Institute of New Media and Communication, Seoul National University, Seoul, Republic of Korea*)

✉ E-mail: lsh529@snu.ac.kr

## References

1 Provos, N., and Honeyman, P.: 'Hide and seek: an introduction to steganography', *Secur. Priv.*, 2003, **1.3**, pp. 32–44
2 Cheddad, A., Condell, J., Curran, K., *et al.*: 'Digital image steganography: survey and analysis of current methods', *Signal Process.*, 2010, **90**, pp. 727–752
3 Subhedar, M.S., and Mankar, V.H.: 'Current status and key issues in image steganography: a survey', *Comput. Sci. Rev.*, 2014, **13-14**, pp. 95–113, doi: 10.1016/j.cosrev.2014.09.001
4 Upham, D.: 'Steganographic algorithm jsteg'. Available at http://zooid.org/~paul/crypto/jsteg(1993)
5 Andreas, W.: 'F5-a steganographic algorithm'. Int. Workshop on Information Hiding, Pittsburgh, PA, USA, 2001
6 Chan, C.-K., and Cheng, L.M.: 'Hiding data in images by simple LSB substitution', *Pattern Recognit.*, 2004, **37**, pp. 469–474
7 Wang, R.-Z., Lin, C.-F., and Lin, J.-C.: 'Image hiding by optimal LSB substitution and genetic algorithm', *Pattern Recognit.*, 2001, **34**, pp. 671–683
8 Kim, C., Lee, S.H., Park, H., *et al.*: 'Image steganography using random permutation and image difference'. Proc. KOSBE, Jeju, Republic of Korea, 2016, pp. 231–234
9 Wu, H.C., Wu, N.I., Tsai, C.S., *et al.*: 'Image steganographic scheme based on pixel value differencing and LSB replacement methods', *Proc. Vision Image Signal Process.*, 2005, **152**, (5), pp. 611–615
10 Swain, G.: 'A steganographic method combining LSB substitution and PVD in a block'. Proc. Int. Conf. on Computational Modelling and Security, Bengaluru, India, February 2016, pp. 39–44
11 Ker, A.D.: 'A general framework for the structural steganalysis of LSB replacement'. Proc. Int. Workshop on Information Hiding, 2005 (LNCS, 3727), pp. 296–311
12 Fridrich, J., Goljan, M., and Du, R.: 'Detecting LSB steganography in color and gray-scale images', *IEEE Multimedia*, 2001, **8**, (4), pp. 22–28