Is Image Steganography Natural?

Guillermo Sapiro
Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455
guille@ece.umn.edu

June 16, 2002

Steganography is the art of secret communication [1, 2]. Its purpose is to hide the presence of information, using for example images as covers. After embedding the secret message into the cover image, a stego-image is obtained. While steganography algorithms create stego-images that are perceptually natural, we questioned if they are statistically natural [3, 4]. We show that stego-images violate recent models of natural images, and discuss the implications of this both in the art of steganography and in the mathematical modeling of natural images.

We study the statistical behavior of natural stego-images, that is, natural images that are contaminated by hidden information. The method used in this study is the one proposed in [3]. It was shown that the distribution of the area of bilevels of natural images obeys a power law. More precisely, consider an image I whose gray levels are between 0 and N. For an integer k, define the bilevel images $(1 \leq l \leq k)$

$$I_l(i,j) = \begin{cases} 1 & \text{if } (l-1)\frac{N}{k} \leq I(i,j) < l\frac{N}{k} \\ 0 & \text{otherwise.} \end{cases}$$

Compute then the number $f(a)$ of connected components of the bilevel images I_l with area a. It was experimentally found and theoretically justified that

$$f(a) \approx \frac{C}{a^\alpha}.$$

In other words, the distribution of areas of the connected components of natural images obeys a power law, or a linear law in log-log coordinates; see Figure 1. Based on simple arguments of scale-invariance, which are fundamental in the modeling of natural images [3, 4, 5, 6], it can be shown that the exponent α must be equal to 2. This is confirmed experimentally.

We proceed to investigate if embedding hidden information violates the above statistics on natural images. We used the popular package S-Tools [1] to create the stego-image, while the same results were obtained with other steganography techniques. Figure 1 shows a representative example of the obtained results. Although the stego-image is still perceptually natural, it violates the power law and statistics of natural images.

These results have implications both in steganography and in the study of the statistics of natural images. Being a fundamental problem the detection of images that contain hidden information, from the results here reported we conclude that we can use recent advanced models of natural images to detect stego-images. This makes the task of steganography much more difficult, since it has to hide the information in a way that preserves sophisticated image statistics. From the point of view of modeling natural images, a good model has to be able to detect stego-images, since these are not natural. This opens an interesting question, the study of the possibilities to imitate the statistics of natural images while embedding coherent information.

Acknowledgments

This work was partially supported by a grant from the Office of Naval Research ONR-N00014-97-1-0509, the Office of Naval Research Young Investigator Award, the Presidential Early Career Awards for Scientists and Engineers (PECASE), a National Science Foundation CAREER Award, and by the National Science Foundation Learning and Intelligent Systems Program (LIS). We acknowledge useful conversations on the topic discussed in this paper with Prof. V. Caselles, Dr. W. Masters, Dr. C. Schwartz, and Dr. G. Seroussi.

References

[1] S. Katzenbeisser and F. A. P. Petitcolas (Editors), *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, 2000.

[2] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color and gray-scale images," *IEEE Multimedia* **8:4**, pp. 22-28, Oct.-Dec. 2001.

[3] Y. Gousseau and J. M. Morel, "Are natural images of bounded variation?," *SIAM Journal on Mathematical Analysis* **33:3**, pp. 634-648, 2001.

[4] D. Mumford and B. Gidas, "Stochastic models for generic images," *Technical Report, Division of Applied Mathematics, Brown University*, 2000.

[5] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Optical Society of America A* **4**, pp. 2379-2394, 1987.

[6] D. L. Ruderman, "Origins of scaling in natural images," *Vision Research* **37:23**, pp. 3385-3398, 1997.

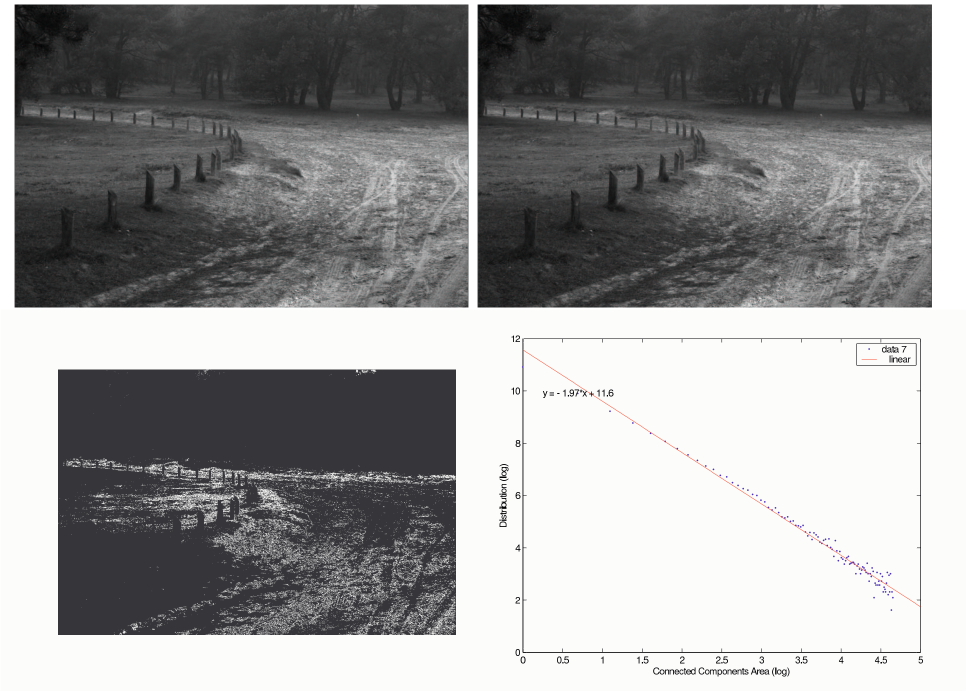

Figure 1: Area density of natural and stego-images. *Top-left*: Original cover natural image (size 1536 × 1024). *Top-right*: Natural stego-image, with hidden data (272 KB). Note how the image is perceptually indistinguishable from the natural cover image, and is perceptually natural. *Bottom-left*: Example of bilevel image of connected components for the original image ($k = 16$ and $l = 7$). *Bottom-right*: Density of the connected components for the natural image. This is a log-log plot, and according to the power law, a linear fit should hold. For the natural image, $\alpha = 1.97$ and $C = 11.6$, which is in the range observed in [3]. For the date base of natural images from van Hateren (hlab.phys.rug.nl/imlib), the average α found for $k = 16$ is 1.85, standard deviation is 0.19, and maximal value of 2.2 (the larger values correspond to textured images). For the natural stego-image, $\alpha = 2.42$ and $C = 13.0$, which are outside of the observed range for natural images. The residual increases from 2.3 to 3.2, showing that the linear fitting for the stego-image is not as accurate. These numbers are typical for the natural cover images, steganography algorithms, and hidden data tested. An increase in the values of α (as well as C and the residual) is observed, shifting it to the tail of the distribution observed for natural images, or even outside the range as in the above example. This shows that although the image with the hidden information is perceptually natural, "statistically" it is not. It also shows that the stego-image violates the scale-invariant property of natural images. Current steganography techniques do not explicitly attempt to preserve this and other important properties of natural images.