Some Properties of the Steganographic Capacity of Images Derived from Watermarking Benchmarks

Klaus Hansen
DIKU
Jens Damgaard Andersen
DIKU
Universitetsparken 1, DK-2100 Copenhagen

June 29, 1999

Abstract

Hiding data in images in an imperceptible but robust way relies on choosing the locations for hiding according to psychophysical measures. The locations may be represented as an image expressing the steganographic capacity in the different areas. i.e. how much the area may be changed without changing the visual appearance of it. Only a few algorithms for computing capacity are found in the literature. In this paper we discuss some properties of algorithms derived from the benchmarks for watermarking proposed in [7] and implemented in the StirMark software. Furthermore the paper proposes to enlarge the set with additional images, with little (or no) capacity, and with unknown capacity.

1 Introduction

Papers on data hiding in images and specifically hiding watermarks in images often implicitly assume that the images in question actually can contain the necessary data, and focus on e.g. robustness or low visibility. This assumption seems reasonable enough for most 'real' images, but still leaves the question of where best to hide data in a given image.

Ross Anderson [1] states that data may be hidden if the entropies of the data to be hidden W and of the host image I meet the requirement $H(W) < H(I)$. Even if this requirement is sensible, it is not operational for several reasons. First, what exactly is meant by 'may be hidden' in terms of the resulting quality is not defined. Second, the entropy $H(I)$ is not well-defined, and any computations beyond first-order statistics meet difficulties. Third, if we aim at invisible or just noticeable hiding, we cannot use a global measure, as different regions of I will hide data with differing success. We need a localised measure of steganographic capacity (SC) for the different regions, see e.g. [6] and [11] . The notion of capacity is not simple, as it depends on the e.g. the planned level of resistance against attacks, on whether the data added may be invisible or not, and on the use of the data.

The amount of data to be hidden depends on the application. For watermarking purposes the amount seems to be a modest 30–160 bit (see [7] or [9]). If cryptographic techniques are used e.g. for authentication, at least 400-500 bit will be needed. If the capacity of a given image is sufficiently large, the resistance to attacks may be increased by using the locations with capacity yet unused to hide random data.

The paper is organised as follows: first (in section 2) we present some general properties of data hiding; then (in section 3) we derive some properties of algorithms for computing capacity from the watermarking benchmarks (see [7]); and finally (in section 4) we propose an extension to the benchmark image set to cover cases with little (or no) capacity, and with unknown capacity.

2 Steganographic capacity of still images

In the following, the data we want to hide are called W, the host image we modify to hide W is called I, and the result of this is $I' = f(I, W)$. The fundamental requirements for a data hiding technique could be that the change should be *invisible*, i.e. that looking at I' or comparing I and I' does not reveal that W is hidden in I', or it may be less strict, meaning that a close inspection of I' will be needed to detect W as the difference is beneath some cognitive threshold for the human visual system, i.e. *just noticeable*. Finally the change may be allowed to be *visible*, if it does not distort I too much. The ITU recommendation R.500 have five quality ratings (see [7]), where the invisible or imperceptible distortion is rated 5 and the just noticeable is rated good or 4. The remaining rates goes from fair quality to bad.

Let us assume that $SC(I)$ is a matrix that for each pixel coordinate gives a measure of the capacity[1] in a small neighbourhood. It may be zero, meaning that any change will be clearly visible. The capacity is subjective, because it depends on sensitivity and masking effects of the human visual system. Some expectations to capacity measure may be formed by looking at e.g. Brodatz textures [4] and could be:

- segments with small variations in grey value have no or small capacity.
- areas along sharp edges have small capacity, areas along blurred edges have larger capacity.
- areas with fine-scale or regular texture have smaller capacity than areas with coarse or more irregular texture[2].
- segments with noise have large capacity.

In the following, it is assumed that the data to be hidden has no distinguishable patterns; this is the *no–pattern condition*. If that is not the case, the possibility exists that patterns in the data will give visible effects which are not revealed by the statistical properties of the resulting image, e.g. as computed by benchmark evaluation software. We will assume that any existing patterns have been removed by e.g. encryption or scrambling with a random pattern, so that the properties of the hiding method do not depend on the data to be hidden.

2.1 Robustness criteria

The level of robustness which is required for a useful algorithm for hiding data does depend on the actual application. Thus a single benchmark suite will not be sufficient for evaluating the performance of all data hiding methods. Invisible watermarks to be used as evidence of origin and ownership are required to be much more robust than data used to identify origin, and how robust simple data hiding is to be depends to a very high degree on the actual circumstances. In this section we try to classify methods with respect to robustness against detection.

We distinguish between *automatic* detection, using some algorithm to detect that data is hidden, and detection by a *human expert* who has been trained to analyse the class of images in question, e.g. a photo reconnaissance image interpreter, who consequently can spot changes that algorithms or *untrained humans* will not detect. By state–of–the–art algorithm is at the moment thought of algorithms using the image statistics or texture models like Markov Random Fields. The classification is:

1. Given one image, possibly with hidden data; determination of presence of data (the *basic* case) is:
   (a) not possible
   (b) possible for human expert, but not feasible for best state-of-the-art algorithms

[1] A property of $SC()$ which makes decoding self–synchronising is that $SC(I')$ is a close approximation to $SC(I)$.
[2] Hiding data by coding in dithering patterns may exploit this

    (c) possible by algorithm
    (d) may be done by close inspection by non-expert
    (e) obvious

2. Given image, possibly with hidden data, and the hidden data W; determine presence of data (the *public mark* case)
    (a) not possible
    (b) not feasible for best state-of-the-art algorithms
    (c) possible by algorithm

3. Given two images, one without I and one possibly with hidden data I' (the *image difference* case); determination of which image is I' is
    (a) not possible
    (b) possible for human expert, but not feasible for best state-of-the-art algorithms
    (c) possible by algorithm
    (d) may be done by close inspection by non-expert
    (e) obvious

Intuitively it is to be expected that hiding data is more difficult with increasing major case number, and that within each case the first item is more difficult than the last. This could be used when evaluating benchmark results to classify a method. The highest classification would be 3a.

In [7] is listed several types of watermarking systems. Their private system corresponds to 3, their semi–private to 1c, and their public to 2c.

2.2 Use of data hiding and attacks

Another classification, which could be used when evaluating benchmark results, is what data hiding is used for.

1. presence and authenticity of data in an image is provable by anyone given algorithm and a public key (e.g. self-enforcing proof of ownership or origin)
2. presence and authenticity of data in an image is provable by trusted third party given algorithm and a key (e.g. proof of ownership or origin)
3. presence of data in an image is provable by anyone, and authenticity is guaranteed by trusted third party (e.g. identification of owner)
4. presence of data is provable by anyone (e.g. indication of owner or origin)
5. private data

The cases are sorted, so the most demanding and difficult used are listed first. Case 1 corresponds to the semi-private type in [7], and case 2 to the public.

Yet another classification scheme is concerned with the attacks a hiding method is designed to withstand. A benchmark indirectly measures resistance to such attacks:

1. Unauthorised detection of hidden data
2. Deletion of hidden data
3. Modification of hidden data
4. Adding new data or interfering with existing data

3 Benchmarking and capacity measures

The benchmarks found in [7] and [8] and implemented in StirMark 3.1 are based on a number of attacks including compression, geometrical transformations and image enhancements techniques. Data hiding should ideally be robust, i.e. resistant to some or all of these attacks, which has the consequence that the corresponding steganographic capacity measure has to have certain properties in order to avoid areas vulnerable to these attacks. The following can be derived:

- geometric transformations: the measure should be independent of coordinate system and pixel position; the measure should have rotation invariance. Resistance to scaling and geometric distortions is reflected in the requirement that the regions are not too small.
- cropping, flipping, distortions: the measure should be non-global covering a neighbourhood of modest size.
- enhancements: the measure should have invariance for bias (average pixel value) and variance (deviation from average)
- compression, deletions, random, small distortions and disturbances: this is not a matter of measuring capacity, but rather of how to code single bit and a bit stream. The techniques called for includes error-correcting codes and redundancy obtained by e.g. spread-spectrum techniques, and by choosing images with a sufficiently large Signal–to–noise ratio SNR.

In [6] we have investigated three capacity measures. One is based on multi-scale analysis. If the difference between the image at one scale and a coarser scale is insignificant, it is likely that changes will be detectable, as we have a homogeneous region. The capacity measure is the difference between an image on the finest level and the next level. The image is thresholded with a suitable value to get a mask. This method uses small neighbourhoods, which means that it is not resistant to distortions, but has otherwise the required properties. Because of the small neighbourhood too many points are given capacity. Including coarser scale results is expected to give a better capacity measure.

Another possible measure in [6] for the capacity is an estimate of the entropy of an 11×11 block surrounding each pixel. It is estimated by counting the number of occurrences of MRF11 patterns in the block[3] and using these as estimates of their probabilities. The measure is based on the corresponding conventional entropy e, being 2^e; the values are thresholded to get a mask. It has the required properties, but the results in the paper as to general usefulness are not convincing, partly due to the limited sample size (121) used to estimate the individual MRF pattern probabilities.

A third possible measure is found in [10] and is based on the energy of the coefficients obtained by an 8×8 Discrete Cosine Transform (DCT). The capacity of one block is the number of coefficients F_{ij} for which

$$|G_{ij}|^2 \geq \varepsilon \sum_{i=0}^{N} \sum_{j=0}^{N} |G_{ij}|^2 \tag{1}$$

where ε is a threshold, which in this paper has the value of 0.2 suggested by [10], and $G_{ij} = W_{ij} F_{ij}$; W_{ij} is a subjective measure of a coefficients visual perceptibility. This methods has the required properties, but as it uses the same blocks as JPEG it is not at all robust against compression.

4 Extending the set of benchmark images

The benchmark set from [8] covers a wide range of image types, but contains no extreme cases, i.e. cases where the hiding either must fail, or it visibly distorts the image. In these cases the hiding algorithm ought to fail instead of hiding data, and application of e.g. StirMark to evaluate the robustness is clearly redundant. It is important that such cases are incorporated in a benchmark.

[3] An MRF11 pattern consists of the pixel and its immediate four neighbours.

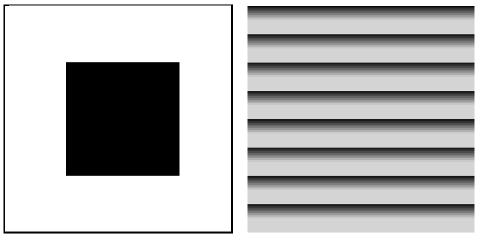

Figure 1: Images with little or no steganographic capacity. Left black and white image base01, right image test11

The steganographic capacity for most images in the benchmark set would be expected to be more than sufficient to hide small amounts of data.

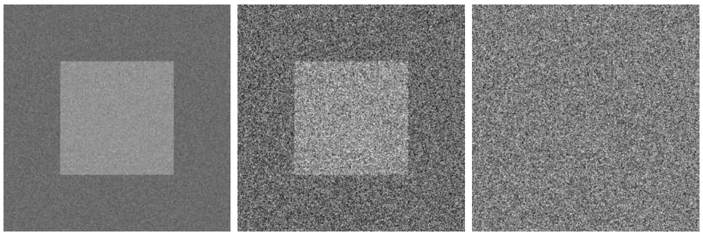

Figure 2: Images with some steganographic capacity due to Gaussian noise. Left S/N ratio 4:1, middle S/N 1:1, right pure noise, std. dev. 40

The two images in figure 1 have little or no capacity, because only very small changes are permissible without clearly visible result. The image 'base01' will even cause some JPEG compression software to give visible distortions even using a modest compression factor.

The three images in figure 2 is expected to have increasing capacity; especially the third image, which contains pure Gaussian noise should be well suited to hide data. The benchmark images ([8], see figure 3) "Localised corrosion ...", "Wildflowers" and "USC texture mosaic #1" belongs to the same class of images. The images "Fourvire Cathedral", "Intergranular Stress Corrosion ...", "Arctic Hare" and "Cleavage of Intermetallics ..." would be expected to have significantly lower capacity.

Figure 3: Four images from [8]: from left to right "Localised corrosion on an electropolished Al-Zn-Mg-Cu alloy", "Arctic Hare", "Fourvire Cathedral, north wall" and "Cleavage of Intermetallics and Hydrogen Embrittlement in an Al-Zn-Mg-Cu alloy". For copyright see [8]. The images are originally colour images

The three images in figure 4 have a capacity for hiding which is unknown.

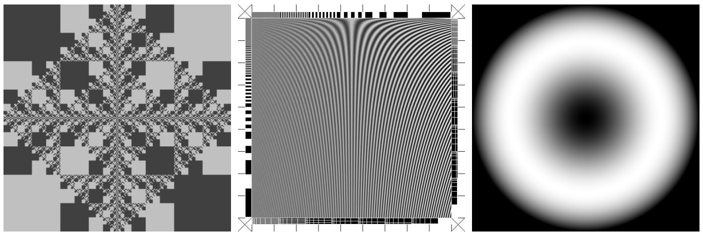

Figure 4: Images with unknown capacity. Left fractal pattern small02, middle sine01 with varying amplitude and frequency, right smooth04, a truncated third degree 'surface'

5 Conclusion

The nature of an image determines the steganographic capacity. Image areas where the precise configuration of primitive image elements or texels does not matter, e.g. in textured areas, have higher capacity than regular areas with relatively uniform gray level. It is not difficult to construct measures of capacity that performs well with the benchmarks, but it seems not easy to design a capacity measure that also takes into consideration the characteristics of pattern detection part of the human visual system. Future research should investigate human image interpretation.

6 Literature

References

[1] Ross Anderson: *Stretching the Limits of Steganography*, in [2].

[2] *Information Hiding* (ed. Ross Anderson) First International Workshop, Cambridge 1996 Springer LNCS 1174.

[3] Ross Anderson: *Information Hiding — an annotated bibliography*.

[4] Phil Brodatz: *Textures*, Dover Publications 1966.

[5] Klaus Hansen: *Test image collection* (http://www.diku.dk/users/khan/pub/image-index.html)

[6] Klaus Hansen and Jens Damgaard Andersen: *Image steganography using spread spectrum techniques* Proc. SCIA 99.

[7] M. Kutter and F. A. P. Petitcolas: *A fair benchmark for image watermarking systems* Electronic Imaging '99, San Jose January 1999

[8] F. A. P. Petitcolas: *Digital Watermarking, Evaluation and Benchmark* (http://www.cl.cam.ac.uk/~fapp2/watermarking/benchmark/)

[9] Laurent Piron et al.: *OCTALIS benchmarking: Comparison of four watermarking techniques* Security and Watermarking of Multimedia Contents, vol. 3657, San Jose January 1999

[10] J.J.K Ó Ruanaidh, W.J. Dowling and F.M Boland: *Watermarking Digital Images for Copyright Protection.*

[11] Sergio D. Servetto, Christine I. Podilchuk and Kannan Ramchandran: *Capacity Issues in Digital Watermarking* Proc. IEEE Int'l Conf. on Image Processing 1998