Steganographic Capacity of Images, based on Image Equivalence Classes

Klaus Hansen
Department of Computer Science
University of Copenhagen
Universitetsparken 1
DK–2100 Copenhagen Ø, Denmark
khan@diku.dk

Christian Hammer
and
Lars R. Randleff
Department of Computer Science
lars@randleff.dk

Jens D. Andersen
Department of Computer Science
jda@diku.dk

ABSTRACT

The problem of hiding information imperceptibly can be formulated as the problem of determining if a given image is a member of a sufficiently large equivalence class of images which to the Human Visual System appears to be the same image. This makes it possible to replace the given image with a modified image similar in appearance but carrying imperceptibly coded information. This paper presents a framework and an experimental algorithm to estimate upper bounds for the size of an equivalence class.

Categories and Subject Descriptors

E.4 [**Data**]: Coding and Information Theory; I.4.6 [**Computing Methodologies**]: Image Processing and Computer Vision

General Terms

Experimentation, Measurement, Security, Theory

Keywords

Information hiding, image equivalence classes, perceptual tolerance, capacity, digital watermarking

1. INTRODUCTION

Digital multimedia have made watermarking an issue of central importance. The degree to which a given image or audio/video data may be modified in an imperceptible way, i.e. the capacity for hiding information, is a topic still being investigated. We suggest an approach to determine if a watermark can be embedded in an image. The method gives an estimate for the upper bound on capacity, but do not tell if a robust and tamper-proof or fragile watermark may be hidden or not. That depends on the coding scheme and the amount of redundancy used. By computing the number of bits used the estimate of the capacity can be used to decide if hiding is viable. The method exploits perceptual properties of the human visual system (HVS) in an experimentally based analysis of the image in question, based on a novel use of the principles utilized in multimedia compression schemes like MPEG-4 ([8]). This paper does not consider how these principles may be used in connection with audio or video data.

We use an image source model resulting in equivalence classes each consisting of perceptually indistinguishable images, in which an image instance is produced by a three-layer process, one giving the high-level structure, one producing middle-level texture, and one representing the low-level noise coming e.g. from the imaging process (A/D, sampling and quantization).

2. IMAGE MODELS AND EQUIVALENCE CLASSES

Ross Anderson has suggested to use Shannon entropy as a measure of the steganographic capacity. The capacity is determined from the entropies of the information to be hidden I and the host image B: information may be hidden giving image B^* when $H(I) < H(B)$ [1]. This criterion is however not very useful. First, a definition of the entropy of a still image is not straightforward because of the lack of appropriate models for the information source. Second, it does not take into account that images are to be viewed, or that automated image interpretation algorithms are designed to imitate the HVS. As entropy is a statistical measure, modifications may leave the entropy unaffected but still be highly noticeable by the HVS.

We assume in this paper that we are given a particular class of images (e.g. natural scenes, aerial photographs, medical images) each described partly by deterministic constraints, partly by stochastic properties. In recent years various research groups have worked on the problem of statistical and information-theoretical characterization of images. There are some results concerning the combinatorial entropy of images (2D fields) [2, 3, 6] and the Shannon capacity in two dimensions. Another line of research is an attempt to construct grammars for visual or two-dimensional languages. A third research direction aims at characterizing images in

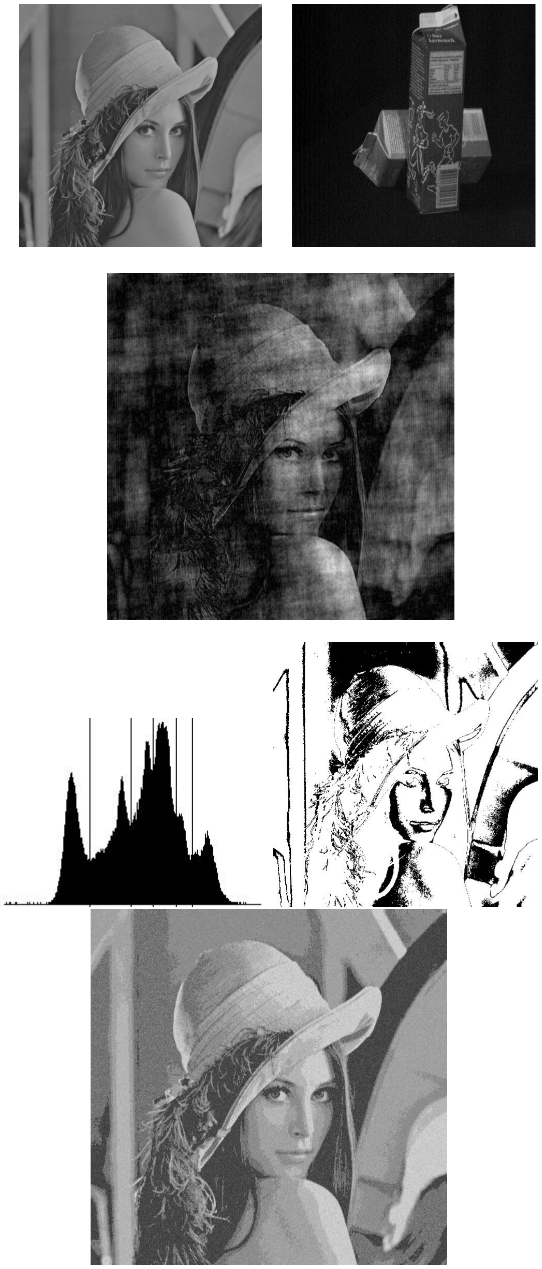

Figure 1: Equivalence classes. (a) top left "Lena" image, right image "milk3"; below the amplitude component of Lena has been replaced by the amplitude from milk3. The structure survives although the image is visibly distorted. (b) left the histogram of Lena divided into six partitions. Right the pixels in the third partition with grey values in the range 128..151. Bottom all pixel values in each partition have been replaced by values taken from a Gaussian distribution with same mean and variance.

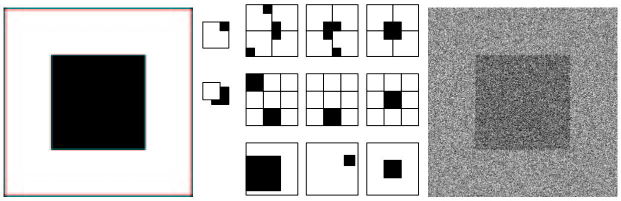

Figure 2: Overall structure and noise. (a) Left an image without capacity if interpreted as one black square on a white background. (b) Middle three of several possible grammars (structure S) generating a square inside another square. Black and white indicates which texture model ($T|S_i$) and stochastic distribution a given pixel will come from. Top a single token having 16 possible configurations. Middle two tokens in nine locations (2^9 configurations). Bottom a square placed anywhere, of any size that fits. With these interpretations of (a), C_S is respectively 4, 9 or the order of M^3 bit. (c) Right noise N has been added. The image in (a) may then be interpreted as a (very unlikely) member of a class having its structure S from (b), uniform texture T and additive noise N distributed as in (c), but by coincidence having zero magnitude.

terms of their multi-fractal properties [9]. Wu et al. [10] has defined Julesz ensembles as equivalence classes of images sharing some statistical properties and consequentially looking similar.

The idea of *equivalence classes* is illustrated in figure 1. The image in the second row is a combination of phase from the left top image and magnitude taken from the right top image. The resulting image belong to the same equivalence class as the left image if we accept poor quality. The phase information seems to be more important than the magnitude information in distinguishing classes. The bottom image shows the results of systematically replacing pixel values within six spatial images partitions. The images top/left and bottom are quite similar.

The problem of hiding information imperceptibly can be formulated as the problem of determining if a given image is a *member of a sufficiently large equivalence class.* If this is the case, it is possible to modify it by a suitable coding scheme. The resulting image has similar appearance but is carrying hidden information, which later may be extracted again. The actual coding of the information to be hidden is similar to modulation in a noisy channel. We assume some parameterization of the images within a class, and code by modifying the parameters systematically. Selecting an appropriate number n of "canonical" images with sufficiently large decoding distance realizes the robustness required.

Let $\mathcal{U}_{M_1,M_2,P}$ be the universe of all possible images of dimension $M_1 \cdot M_2$ consisting of P-bit pixels. The size $|\mathcal{U}_{M_1,M_2,P}|$ is $2^{M_1 M_2 P}$. Only a very small fraction of this huge number of images is of any interest (carries true pictorial information representing scenes from a model universe $\aleph$). An image $I_{M_1,M_2,P} \in \mathcal{U}_{M_1,M_2,P}$ has some structure S (described by e.g. a 2D grammar), texture T (which may be parameterized using a Markov model and locally depend on S_i, notation $T|S_i$) and a stochastic element N, independent of T and S. See figures 2 and 1(b).

If $\Omega_{M_1,M_2,P}(S) \subseteq \mathcal{U}_{M_1,M_2,P}$ is the collection of images with same overall structure S, then $|\Omega_{M_1,M_2,P}(S)| = K_S$. We denote by e.g. K_S and $K_{T|S_i}$ the size of equivalence classes relating to S and $T|S_i$. $K_S = \prod_i K_{T|S_i}$, as the texture depends on the actual local structural component S_i. As the noise may be assumed independent ($K_{N|T,S_i} = K_N$), $K_{T|S_i} = K_N K_{T|N,S_i}$. When convenient we will use $C_S = \log_2(K_S)$ instead of K_S etc., and $A_S = \frac{C_S}{M_1 \cdot M_2}$, the average number of bits per pixel. C_S may be interpreted as the image entropy $H(I)$.

Robust encoding of information in one of the components is possible if the corresponding K is large enough. Estimation of these is not straightforward. Determining the number and size of equivalence classes involves the given model universe $\aleph$ and properties of the human visual system. Using the structural element S seems difficult, as the number of classes will be rather small and concerns the overall spatial properties of an image. Modifying the least significant bit of each pixel is based on the assumption that $K_S = K_N = M_1 \cdot M_2$ bit.

From the six normal distributions of the histogram of Lena in figure 1 we can compute the contributions to the $K_{T|S_i}$ and K_N from the variances. Of the approximately 7.5 bit/pixel (entropy derived from the pixel frequencies) the variance contributes at least 4.5 bit/pixel (average of the variances for each of the six partitions). A rough guess of the real noise level (N) is 30 dB or 0.2 bit/pixel, leaving 4.3 bit/pixel for texture (T). Assuming that the structural part S contributes 0.1 bit/pixel, we may compute $A_S = 4.3$ bit/pixel. The hiding capacity C_S for Lena is then $512 \cdot 512 \cdot 2^{4.3}$ or 5 163 793 bit. This does not take into account the properties of HVS, but gives an absolute upper bound. By the method described in section 4 a capacity of 1,382,729 bit is found for a 480 × 360 × 24 bit version of Lena. For some images K_T will be insignificant, as illustrated in figure 2, where all capacity comes from S, C_S = 4–9 bit.

The following sections of this paper provides a method for determining $K_{T|S_i}$ and the distribution of capacity over an image based on properties of the HVS.

3. PERCEPTUAL TOLERANCE

We define the *perceptual tolerance* at a pixel coordinate in an image P as the amount that each pixel component (e.g. RGB, or HSV) can be changed without changing in the perception. The total tolerance is the product of all pixel tolerances. Images having pixel values in the tolerance intervals belongs to the same equivalence class. The *capacity* for hiding information is not expected to fully exploit the tolerance.

The perceptual tolerance may be determined experimentally by combining a number of filter responses (here F_i for i in $1 \dots 6$) each enlarging the tolerance interval. For each pixel component $p_{ijk} \in P$ (i, j being the pixel coordinates, k indicating the component) the tolerance is the interval from $\underline{p}_{ijk}$ to $\overline{p}_{ijk}$. These values are stored in the images $\underline{P}$ and $\overline{P}$. The algorithm has the following steps:

1. Initially $\underline{P} = \overline{P} = P$.
2. For each filter index i:
   (a) Given P and a filter F_i an image $P^* = F_i(P)$ is generated. F_i is chosen in such a way that there will be no perceptual difference between P and P^*.
   (b) $\underline{P}$ and $\overline{P}$ are updated by $\underline{P} = min(\underline{P}, P^*)$ and $\overline{P} = max(\overline{P}, P^*)$.
3. The resulting tolerance (the number of images in the equivalence class) is $K_S = \prod_{ijk}(\overline{p}_{ijk} - \underline{p}_{ijk} + 1)$, of the tolerance at each position.

Steps 2a and 2b are repeated for filters that can be related to certain vision phenomena, as identified in [5]. Descriptions of these phenomena can be found in e.g. [4] and [7]. Figure 3 shows the tolerances in each pixel position of an image of Big Ben, brighter pixels having larger tolerance.

The filters used concerns $K_{T|S_i}$:

- Mono- and polychromatic assimilation. According to [4] HVS makes thin lines the same intensity as surrounding areas. This monochromatic assimilation is estimated by threshold based mean filters $F_1(P_{assim,m})$ (equal weigth) and $F_2(P_{assim,w})$ (weighted). A corresponding use of filters can probably be used for polychromatic assimilation.
- Area homogeneity $F_3(P_{H_v})$ (value) and $F_4(P_{H_h})$ (hue). In a 3 × 3 pixels area the variances for hue and value are determined and thresholded.
- Just noticeable distortion JND. A JPEG-compression with a suitable loss factor P_{jnd} has been used as $F_5(P_{jnd})$, the difference when decompressing again is used as a measure.
- Edge enhancement. The position of edge pixels was determined by a simple 3 × 3 Sobel filter applied to the *value* component, and $F_6(P_{edge})$ is determined as P_{edge} percent of the edge strength found.

4. TOLERANCE RESULTS

In [5] the applicability of the filter set was evaluated by asking 18 subjects. Each subject looked at 106 sets of nine images. Either the nine images were identical, or two of them (identical) had been modified. A subject was asked to identify zero, one or two differing images. The aim of the study was to determine the perceptual tolerance and the values for the filter parameters. The 106 sets of images were based on six basis images.

The parameter values to be used are those causing at most half of the subjects to notice any difference. The results of the study for two of these parameter values are shown in figure 4. The upper plot illustrates the determination of p_{edge}. Two graphs are shown since each parameter value was tested in two different ways. The lower plot in 4 shows the 3D plot for p_{H_v}, where the abscissae shows repectively the threshold and the percentage change in pixels chosen by the threshold used.

Our hypothesis is that the more a given parameter allowed an image to be changed, the easier it is for the subject to notice the changes. This hypothesis was proven to be true for some of the parameters involved.

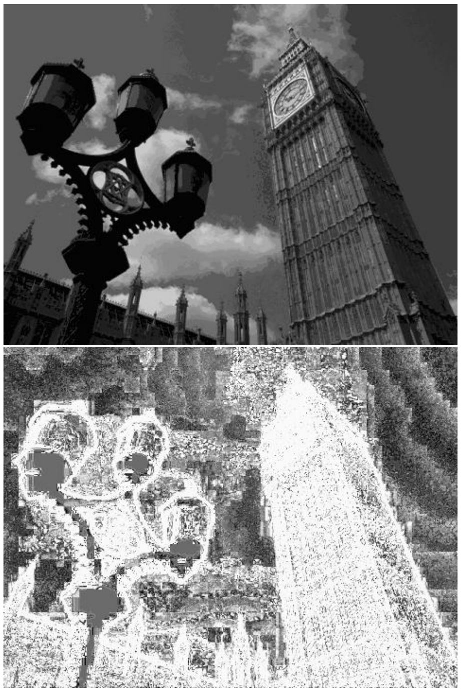

Figure 3: Tolerance. Top the 480 x 360 image "Bigben", bottom the tolerance image (the difference between the bounding images), dark signifying low tolerance, white high tolerance.

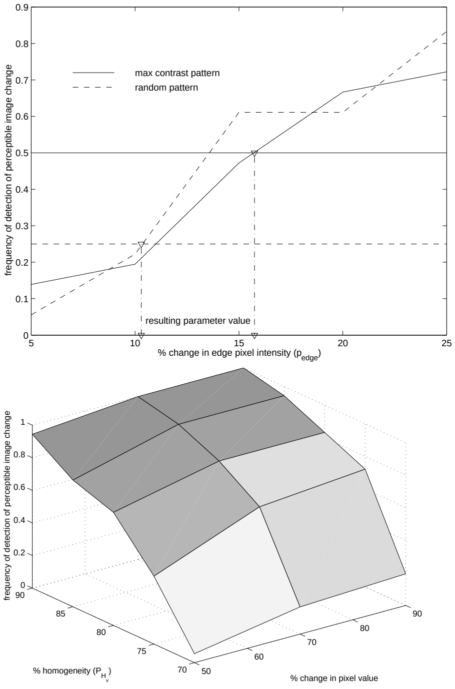

Figure 4: Test results. (a) top, the influence of increasing the modifications near an edge (visibility versus P_{edge}). (b) bottom, the results of changing pixels in homogeneous areas (visibility for random pattern versus P_{H_v} and resulting pixel change).

Using the parameter values found we were able to calculate the perceptual tolerance in a number of images. The corresponding total capacity is 1,283,542 bit for Bigben and 1,382,729 bit for Lena.

5. DISCUSSION AND CONCLUSION

This article presents a model which gives a basis for determining whether information hiding is possible, and if possible, then to which extent in each individual image.

Known psychophysical properties of the human visual system are expressed as filters and tolerance limits, which makes it possible to determine an upper bound on the capacity of an image. The actual capacity depends on how the information is coded and the redundancy requirements.

6. ACKNOWLEDGEMENTS

This work have partly been supported by the Danish Natural Science Research Council, grant no. 21–00–0370.

7. REFERENCES

[1] R. Anderson. Stretching the limit of steganography. In *Information Hiding, First International Workshop (Springer Lecture Notes in Computer Science 1174)*, 1996.

[2] R. Buccigross and E. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8(12):1688–1701, 1999.

[3] J. de Bonet and P. Viola. A non-parametric multi-scale statistical model for natural images. *Advances in Neural Information Processing*, 10, 1997.

[4] A. Fiorentini, G. Baumgarten, S. Magnussen, P. H. Schiller, and J. P. Thomas. *The Perception of Brightness And Darkness, Visual Perception: The Neurophysical Foundations.* Academic Press, Inc., 1990.

[5] C. Hammer and L. R. Randleff. Steganografi i billeder baseret på en model af human perception. Master's thesis, Dept. of Computer Science, University of Copenhagen, 2000. In Danish.

[6] J. Justesen and Y. M. Shtarkov. The combinatorial entropy of images. *Problemy Peredachi Informatsii*, 33:3–11, 1997.

[7] R. A. Moses and J. William M. Hart. *Physiology of the eye, Clinical Application.* The C.V. Mosby Company, 1987.

[8] M.-T. Sun and A. B. Reibman, editors. *Compressed Video over Networks.* Marcel Dekker, 2001.

[9] A. Turiel and N. Parga. The multi-fractal structure of contrast changes in natural images: from sharp edges to textures. *Neural Computation*, 12:763–793, 2000.

[10] Y. Wu, S. Zhu, and X. Liu. Equivalence of Julesz ensembles and FRAME models. *International Journal of Computer Vision*, 38(3):245–261, 2000.