Steganographic embedding in containers-images

To cite this article: A V Nikishova et al 2018 J. Phys.: Conf. Ser. 1015 042041

View the article online for updates and enhancements.

Related content
- Methods of Data Processing for Trace Elements Analysis Using Laser Induced Breakdown Spectroscopy
  Wang Rui, Ma Xiaohong, Yu Qi et al.
- A method of data structuring in the decision-making support system in oncological diagnostics of prostate diseases
  S M Zaytsev, V G Nikitaev, A N Pronichev et al.
- Erratum: Quality assessment of MEG-to-MRI coregistrations (2018 Phys. Med. Biol. 63 075003)
  Hermann Sonntag, Jens Haueisen and Burkhard Maess
Steganographic embedding in containers-images

A V Nikishova, T A Omelchenko, S A Makedonskij
Volgograd State University, 100, Universitetskij Av., Volgograd, 400062, Russia
E-mail: nikishova.arina@volsu.ru

Abstract. Steganography is one of the approaches to ensuring the protection of information transmitted over the network. But a steganographic method should vary depending on a used container. According to statistics, the most widely used containers are images and the most common image format is JPEG. Authors propose a method of data embedding into a frequency area of images in format JPEG 2000. It is proposed to use the method of Benham-Memon-Yeo-Yeung, in which instead of discrete cosine transform, discrete wavelet transform is used. Two requirements for images are formulated. Structure similarity is chosen to obtain quality assessment of data embedding. Experiments confirm that requirements satisfaction allows achieving high quality assessment of data embedding.

1. Introduction
Information security is one of important aspects of digital communication. A necessary step is to protect information before data are transmitted over open communication channels in order to reduce its vulnerability to the many threats typical for this environment. Steganography is one of the approaches to ensuring the protection of information transmitted over the network. Steganography is advancing rapidly nowadays; new approaches are developed.

For example, in article [1], the author proposes a new approach that at first encrypts data and then embeds bits of message into image bits from lowest to highest, depending on the length of the message. To determine the position of embedded bits of the message, there are the secret key sums by modulo 2 with the green component of the image.

In research [2], the author proposes a cost-effective way to hide information about the patient's diagnosis in an electrocardiogram signal based on an intellectual shift factor.

In paper [3], a mechanism using a hybrid detector of lowest bits is proposed. The method is based on hiding in the texture data of the image.

The author of [4] proposes a scheme of data hiding without loss for embedding messages in each block of discrete cosine transform coefficients. Two adjacent coefficients whose values are close to zero are selected as a group in each block of coefficients.

In [5] reversible data hiding a scheme based on secret sharing of Shamir to verify legal ownership are proposed. Information about the container is divided into several pseudo random parts. A secret owner defining information is embedded in encrypted parts of the container based on the private key.

However most part of recent approaches have disadvantages, e.g. they are not resistant to active attacks on steganocontainers. So the model of data steganographic embedding into containers-images is proposed.

The main elements that determine operation of a steganographic system are the container used to hide protected data and the method of protected data embedding in the container.
2. Selection of container and method of data embedding

To increase the invisibility of protected data transfer, the container should be natural for an interaction process in which the steganosystem operates. It is also desirable that the container has a large capacity for embedding information. The most commonly used containers on the Internet are images. According to statistic in figure 1, they also occupy the largest amount of information stored on the sites. Therefore, it is most preferable to use images as steganographic containers.

According to statistic of httparchive.org in figure 2, the most common graphical format is JPEG. However, statistics does not take into account versions of the JPEG format. When developing the software of steganographic embedding, it is proposed to use a modern version of the selected format — JPEG 2000.

An algorithm for converting graphic images in JPEG 2000 consists of several steps performed on the image consistently:

- Color space conversion;
- Underdiscretization;
- Discrete wavelet transform;
- Quantization;
- Encoding.

During the color space conversion phase, the image is converted from RGB color space to YCbCr, where Y is brightness and Cb and Cr are color-difference components of the image point.

During underdiscretization, the phase size of Cr and Cb planes is reduced. The most common reduction is two times in width and two times in height.

Then, separately for each component of color space Y, Cb and Cr, a direct discrete wavelet transform is performed. Use of the discrete wavelet transform allows one to move from spatial representation of the image to spectral representation.

After that, the resulting information is quantized, i.e. some amount of information is discarded. At this stage, the so-called quantization tables are used. They are matrices of 8x8 size consisting of positive integers. Frequencies of image blocks are divided by the corresponding elements of tables.

Encoding is the final stage of compression, during which image blocks are converted to a vector form. Resulting data are then compressed using arithmetic coding or a modification of Huffman's algorithm.

As for methods of data embedding, they are divided into two main classes: methods of direct substitution and spectral methods. The first group of methods uses spatial area, and the second — spectral area. The methods that are more resistant to a variety of distortions, including compression,
are rather those that use spectral area of the container than the ones that use spatial area.

Analysis shows that the best method is spectral methods of Benham-Memon-Yeo-Yeung. However, it uses the discrete cosine transform. The authors propose developing software using methods of Benham-Memon-Yeo-Yeung, where instead of the discrete cosine transform, the discrete wavelet transform is used.

3. Model of steganographic embedding in containers-images

To formalize the steganographic embedding performed in the developed software, the functional model was developed. It is presented in figure 3.

![Figure 3. Detail IDEF0 diagram of software operating.](image)

The input data of software are an image into which data will be embedded, and data themselves that need to be embedded. The functioning of the software is determined by the characteristics of JPEG2000 and algorithm Benham-Memon-Yeo-Yeung. Transformations are performed using developed software and plug-in MatLab library. The result of software is a steganocontainer with data embedded in it, quality assessment, obtained by comparing the steganocontainer with the initial container, and data extracted from the steganocontainer. The functioning of the software complex includes three large functional blocks: embedding, assessing quality and extracting.

The result of the embedding block is a container-image with data embedded in it — steganocontainer.

Input of the quality assessing block is the empty container-image and the steganocontainer formed by the embedding block. For assessing the quality, methods of distortion control are used. Distortion control allows not assessing a qualitative side of the modified container, which depends on a variety of factors, all of which are not possible to take into consideration, but giving a numerical assessment of changes that have occurred as a result of embedding a particular message in a particular container.

The authors propose using an index of structural similarity (SSIM) for quality assessing. The essence of the method is to determine the degree of similarity between two images. A distinctive feature of this criterion is that it gives good approximation to human vision. The index of structural similarity is defined by the formula (1):

\[
SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_x^2 + \mu_y^2 + c_1(\sigma_x^2 + \sigma_y^2 + c_2)},
\]

\[
c_1 = (k_1 L)^2; \quad c_2 = (k_2 L)^2,
\]
where $\mu_x$ – mean of $x$, $\mu_y$ – mean of $y$, $\sigma_{xy}$ – covariance of $x$ and $y$, $\sigma_x^2$ – dispersion of $x$, $\sigma_y^2$ – dispersion of $y$, $x$ – block of original image, $y$ – block of image with embedded message, $k_1$ and $k_2$ constant values, $L$ – maximum value of pixel.

SSIM is a value ranging from -1 to 1, and value 1 is only available in case of two identical blocks of images.

In the embedding block replacing the least significant bit in a sequence of wavelet coefficients obtained after two-dimensional wavelet, the transform of the original image is performed. An algorithm of block embedding is applied.

4. Experiments

As part of experiments, it is necessary to assess how effective the developed software is for different containers-images of format JPEG 2000.

According to algorithm Benham-Memon-Yeo-Yeung, it is proposed that embedding the algorithm used in the software also has two limitations when selecting container blocks for embedding:

- suitable blocks should not have abrupt changes in brightness;
- suitable blocks should not be too monotonous.

On this basis, experiments include two groups of studies:

- in the first group, mainly those images that do not contain blocks that do not meet the requirements will be used. In this case, the software will check units for suitability.
- the second group will use images containing a large number of blocks that do not meet the requirements.

The second group is divided into two subgroups:

- the first subgroup includes images containing a large number of blocks with abrupt changes in brightness;
- the second subgroup includes images containing a large number of too monotonous blocks.

In the second group, the software will use all blocks of the container-image. For both groups, quality assessment of embedding data will be calculated by the developed software. For each group (or subgroup), embedding in 10 empty containers-images is carried out. An example of software functioning during the first group of experiments is shown in figure 4.

![Figure 4. Result of data embedding with quality assessment.](image-url)
Obtained quality assessment for one of the experiments of the first group shown in the figure is high.

5. Conclusion
As a result of experiments, the average value of quality assessment for each group of experiments is obtained and shown in Table 1.

Table 1. Values of quality assessment when embedding in different containers-images.

| Group of containers-images | Average value of quality assessment |
|----------------------------|-------------------------------------|
| Containers-images, mainly not containing blocks that do not meet the requirements | 0.97 |
| Containers-images containing a large number of blocks with abrupt changes in brightness | 0.68 |
| Containers-images containing a large number of too monotonous blocks | 0.52 |

Analysis of experiments’ results has shown that the greatest quality assessment of data embedding with the developed software have empty containers-images mainly not containing blocks that do not meet the requirements. So requirements satisfaction in accordance with the method of Benham-Memon-Yeo-Yeung allows making embedding more unnoticeable and therefore more effective.

6. Acknowledgments
The reported study was funded by the Council for grants of the Russian Federation President according to research project No. MK-6404.2018.9.

References
[1] Seeja K R, Rana J, Priya S and Ahuja L 2018 A novel edge based image steganography technique Advances in Intelligent Systems and Computing 614 66-75
[2] Yang C Y and Wang W F 2018 High-capacity ECG steganography with smart offset coefficients Smart Innovation, Systems and Technologies 81 129-136
[3] Lee C F, Shen J J and Chen Z R 2018 Steganographic image hiding schemes based on edge detection Smart Innovation, Systems and Technologies 81 12-30
[4] Weng C Y, Huang C T and Kao H W 2018 DCT-based compressed image with reversibility using modified quantization Smart Innovation, Systems and Technologies 81 214-221
[5] Singh P and Raman B 2018 Reversible data hiding based on Shamir’s secret sharing for color images over cloud Information Sciences 422 77-97
[6] Akansu A N and Haddad R A 2002 Multiresolution Signal Decomposition: Transforms, Subbands and Wavelets (New York: Academic Press Inc)
[7] Sarmah D K and Kulkarni A J 2018 JPEG based steganography methods using Cohort Intelligence with Cognitive Computing and modified Multi Random Start Local Search optimization algorithms Information Sciences 430-431 378-396
[8] Wu H T, Tang S, Huang J and Shi Y Q 2018 A novel reversible data hiding method with image contrast enhancement Signal Processing: Image Communication 62 64-73
[9] He W, Cai J, Xiong G and Zhou K 2018 Improved reversible data hiding using pixel-based pixel value grouping Optik 157 68-78
[10] Dalmia N and Okade M 2018 Robust first quantization matrix estimation based on filtering of recompression artifacts for non-aligned double compressed JPEG images Signal Processing: Image Communication 61 9-20