Abstract: In this paper we propose a new information hiding and extracting method without embedding any information to a target content using non-linear feature extraction trained on frequency domain. Our system can detect a hidden bit code from the content by processing coefficients of the selected feature block of frequency domain. For the generation of the keys which is needed for extracting a bit code from a content, is done by a supervised learning of the set of values in the selected feature block with the teacher signal value. The teacher signal value is the bit code that you want to relate to the content. The connection weight which was processed by the supervised learning will be used as the key for extracting the bit code that you related to the content. With our proposed method, we were able to introduce a watermark scheme with no damage to a target content because there are no information added to the target content, and this characteristic is effective when you don’t want the target content to be damaged at all.

Key-Words: damage-less embedding, non-linear feature extraction, supervised learning on DCT domain, robust watermark, feature selection on frequency domain

1. Introduction

Recently, because the information technologies made a rapid progress, it became easier to create digital multimedia contents. This situation helped to publicly open up their digital contents even for ordinary personal computer users, for example, distributing through the internet. But the fact that digital contents are easy to make an exact copy helped unintentional and unauthorized distribution of contents through the internet also. Therefore, digital watermark became as a key technology which would protect those copyrights to the author of the contents from unauthorized change of its contents. In general, a digital watermark is a technique to conceal a code imperceptibly to the observer of the content, which is also difficult to remove the code from the content, in order to protect intellectual property rights.

There are other ways to protect a digital content. For example by encrypting the content with a decryption key shared between the author and the observer. But with this method, other observers without a decryption key are not able to access to the content, avoiding a free distribution and circulating widely through the network, which might not be desirable to the author. Where as watermark only embeds data like copyright information which is not needed to be shared among observers, although you can always ask author for it.

Digital watermark is often embedded imperceptibly to human receptors for not disrupting the contents with information which you embed. This is mostly done by transforming multimedia content using frequency transform method such as FFT, DCT and DWT. And data are embedded into the frequency domain.

This paper presents a new model of watermark which would not embed any data to the content and to extract meaningful data from the content by processing feature coefficients using supervised learning. Our model trains neural network to assign predefined secret data and use the neural network weight as a key to extract the predefined code. This feature means that it would not
2. Related works

There are many watermark method make use of the characteristic of a contents itself and/or human sensuous characteristic. The latter have become a major research topic in imperceptible watermark.

Watermark, in general, a small signal just enough for observer to be imperceptible are embedded by a specific process and are able to extract the signal from the content. For those specific process, the model of a human visual system model [2][3] are used for still images and Psycho acoustic model [4][5] are used for sounds. These models imitate complex and elaborate human perceptive system to support for recognizing the patterns and extracting features from an observed object.

According to human visual system, human eyes are less accurate in the regions where the change is rapid, and is sensitive in the regions where the value is changed smoothly. So normally signals are embedded in those areas with extreme alteration or in areas of high frequency domain after image has been transformed using frequency transform method.

In our model, we apply human visual systems but we do not embed any data to the content at all. We only use those feature values for the supervised neural network learning.

3. Proposed method

In this section, we present the hidden value embedding method to the neural network using the selected feature values, and extraction method using the network weight and the selected feature values.

Notice that with our proposed method, we do not embed a bit of data to the content but we make use of the feature values of the content in order to relate or learn the hidden values to the content using back propagation learning. Back propagation learning is a supervised learning, which is one of neural network algorithm for non linear adaptive signal processing system and is a multi layer perceptron model. In detail, for the embedding phase, we select the feature block from the target image and use the feature coefficient values in that block as an input data for the neural network. And it is trained with the teacher values, which is a target values that you want to relate to the selected features. The network weight in that training will be the keys you need for signals to be extracted in the extraction phase.

In case of images that require DCT for the frequency transforms, like JPEG format, the feature values are selected diagonally from the feature block. This selection enables us to pick up the DC coefficient and AC coefficients from high frequency domain to low frequency domain, and these values characterizes most of the features in that selected block. Also, the characteristic of multi-layer perceptron and the training with back propagation makes the system to be robust to the alteration of the feature values. The distribution of the coefficients for DCT are shown below in Fig1.

![Fig1. coefficients in frequency domain](image)

3.1. Multi-layer perceptron and back propagation learning

In this paper, we apply multi-layer perceptron neural network model for adaptive signal processing algorithm and back propagation learning for the creation of the decryption key. The use of multi-layer perceptron mainly
has two advantages. Firstly, multi-layer perceptron can cluster a class non-linearly, so it is able to process the nonlinear behavior. Secondly, it requires less statistical assumptions about the situation where the neural network is in working. We used three layer perceptron network as a model, as Fig2. shown below.

Proposed method is processed by following steps: initialization of the network with proper number of neurons, learning process of the input data by feed-forward computation and feed-back computation of neurons respectively to the teacher signals. For the feed-back learning computation, we use the general delta rule to modify connection weight of each neuron. In this paper, embedding is a learning process for the neurons with teacher values and creation of network weights, which act as the decryption keys. Input and output function that we used are sigmoid function.

### 3.2. Embedding and extracting method

In this section, we discuss the embedding and extracting method for still image, especially on JPEG images in detail.

#### 3.2.1. Embedding and generating extraction keys

First we calculate the DCT coefficients of a image. Normally DCT makes a set of 8 pixels * 8 pixels sub-blocks. Then quantization is processed by using standard JPEG quantization table. Secondly we will choose the DCT sub-block that was selected for the embedding process.

Finally, we will use the network weights that were generated in the embedding process and use those values for the feed-forward network to make a response signal. Then the diagonal coefficients from the selected sub-block are used as an input for the network and result of the feed-forward computation, we will have response of embedded signals as the output of feed-forward network. This output signals are taken through a fixed threshold, and the recognition of the output signals are taken place.

#### 4. Experiment and future works

Results of the preliminary experiment showed that we were able to embed 8 bits of data per one sub-block and was able to respond a different signals distinguishing
between the proper input data and incorrect input data as shown in Fig4. and Fig5. respectively. Embedded data for this experiment were \(1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \) and it clearly shows that the signal was recovered only when the correct input data was used.

Theoretically, we are able to embed more data in a sub-block and this function is still remained to be implemented as our future work. Also if we could define feature values for frequency domain, we can apply the same method in any kind of frequency transforms. This means that we could embed data to even JPEG 2000 formats where its frequency transform method is DWT, only if we can define the feature values of the image.

**References**


5. Conclusion

In this paper, we proposed a new watermark scheme that does not damage the target contents at all. This was done by training a selected feature values of an image to desired hidden data using supervised learning. And the network weights that were generated by the training is saved outside the system and kept by the author. The author then uses that network weights to recover the hidden data by feed-forward computation of the feature values to the network with that network weights.