

# AN EVALUATION OF WAVELET FILTERS PERFORMANCE FOR STEGANALYSIS

Zohaib Khan, Atif Bin Mansoor

National University of Sciences and Technology, Pakistan  
zohaibkh\_27@yahoo.com, atif-cae@nust.edu.pk

## ABSTRACT

This paper presents an evaluation of wavelet filters performance for the task of steganalysis. We analyzed six different wavelet filters namely Daubachies, Coiflets, Symlets, Discrete Meyer, Biorthogonal and Reverse Biorthogonal families for feature extraction in a wavelet based steganalysis technique. Two publicly available steganography tools, namely the F5 steganography and the Model Based steganography were used to embed messages in a database of clean images to develop steganographic database of images. A Fisher Linear Discriminant classifier is trained using all six feature sets extracted from both clean and steganographic images separately and subsequently used for classification. Experiments revealed that the features using the 'haar' (*dbl*) wavelet filter gave the best steganalysis performance.

**Index Terms**— Steganography, Steganalysis, Wavelet filters, Feature extraction, Classification

## 1. INTRODUCTION

The word steganography comes from the Greek words *steganos* and *graphia*, which together mean 'hidden writing'. Steganography is the art of hiding a message in plain sight. In the digital sense, it involves embedding a secret message file into an inconspicuous cover file, such as an image [1].

Steganography is an ancient subject, with its roots lying in ancient Greece and China, where it was already in use thousands of years ago. However, the modern formulation of steganography is often given in terms of the prisoners' problem [2], where Alice and Bob are two accomplices in a jail who wish to communicate in order to hatch an escape plan. However, all communication between them is examined by the warden, Wendy, who will put them in a high security prison at the slightest suspicion of covert communication. Specifically, in the general terms of a steganography model shown in Figure 1, we have Alice wishing to send a secret message  $m$  to Bob. In order to do so, she 'embeds' secret message  $m$  into a cover-object  $c$  according to a shared secret key  $k$  to obtain the stego-object  $s$ . The stego-object  $s$  is then

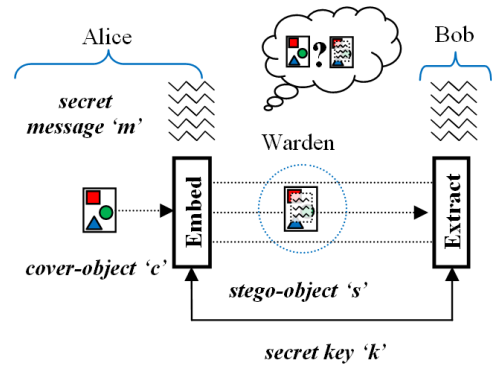


Fig. 1. A general steganography model

sent by Alice through the public channel to Bob,  $m$  unnoticed by Wendy. Once Bob receives the stego-object  $s$ , he is able to recover the secret message  $m$  using the shared secret key  $k$ .

Steganography and cryptography are closely related data hiding methods. The purpose of cryptography is to scramble a message so that it cannot be understood, while that of steganography is to hide the message so that it cannot be seen. In general a message in cipher text might arouse suspicion on an observer while an 'invisible' message created with steganographic methods will not. Sometimes, steganography and cryptography are combined in a way that the message may be encrypted before hiding to provide additional security.

Those who conceal communications through steganography are countered by those who wish to unveil such communications. The field devoted to counter steganography is known as *steganalysis*. The first and foremost goal of a steganalyst is to detect the presence of steganography so that the secret message may be stopped before it is received. Then the second goal is to identify the steganography tool so that the secret message may be spoofed and/or corrupted or even extracted from the stego file.

Generally, two approaches are followed for steganalysis; one is to come up with a steganalysis method specific to a particular steganographic algorithm. The other is to develop universal steganalysis techniques which are independent of the

This work was supported by National University of Sciences and Technology, Pakistan.

steganographic algorithm. Both approaches have their own strengths and weaknesses. A steganalysis technique specific to an embedding method would give very good results when tested only on that embedding method; but might fail on all other steganographic algorithms as in [3], [4], [5] and [6]. On the other hand, a steganalysis technique which is independent of the embedding algorithm might perform less accurately overall but still shows its effectiveness against new and unseen embedding algorithms as in [7], [8], [9] and [10]. Our research work is concentrated on the second approach due to its wide applicability.

In this paper, we propose an evaluation of wavelet filters performance for steganalysis. We extract features using six different wavelet filters which are subsequently used for classification using Fisher Linear Discriminant classifier. The rest of the paper is organized as follows: In Section 2, we discuss the previous research work related to steganalysis. In Section 3, we present our proposed approach. Experimental results are presented in Section 4. Finally, the paper is concluded in Section 5.

## 2. RELATED WORK

Due to the increasing availability of new steganography tools over the internet, there has been an increasing interest in the research for new and improved steganalysis techniques which are able to detect both previously seen and unseen embedding algorithms. A good survey of benchmarking of steganography and steganalysis techniques is given by Kharrazi et al. [11].

Fridrich et al. presented a steganalysis method which can reliably detect messages hidden in JPEG images using the steganography algorithm F5, and also estimate their lengths [3]. This method was further improved by Aboalsamh et al. [4] by determining the optimal value of the message length estimation parameter  $\beta$ . Westfeld and Pfitzmann presented visual and statistical attacks on various steganographic systems including EzStego v2.0b3, Jsteg v4, Steganos v1.5 and S-Tools v4.0, by using an embedding filter and the  $\chi^2$  statistic [5]. A steganalysis scheme specific to the embedding algorithm Outguess is proposed in [6], by making use of the assumption that the embedding of a message in a stego image will be different from embedding the same into a cover image.

Avcibas et al. proposed that the correlation between the bit planes as well as the binary texture characteristics within the bit planes will differ between a stego image and a cover image, thus facilitating steganalysis [7]. Farid suggested that embedding of a message alters the higher order statistics calculated from a multi-scale wavelet decomposition [8]. Particularly, he calculated the first four statistical moments (mean, variance, skewness and kurtosis) of the distribution of wavelet coefficients at different scales and subbands. These features (moments), calculated from both cover and stego images were

then used to train a linear classifier which could distinguish them with a certain success rate. Fridrich showed that a functional obtained from marginal and joint statistics of DCT coefficients will vary between stego and cover images. In particular, a functional such as the global DCT coefficient histogram was calculated for an image and its decompressed, cropped and recompressed versions. Finally the resulting features were obtained as the  $L_1$  norm of the difference between the two. The classifier built with features extracted from both cover and stego images could reliably detect F5, Outguess and Model based steganography techniques [9]. Avcibas et al. used various image quality metrics to compute the distance between a test image and its lowpass filtered versions. Then a classifier built using linear regression showed detection of LSB steganography and various watermarking techniques with a reasonable accuracy [10].

## 3. PROPOSED APPROACH

### 3.1. Feature Extraction

Since the dimensionality of image data is huge, it is not feasible to use the complete image data directly for steganalysis. A better option is to extract a certain amount of useful data and use it to represent the image instead of the image itself for steganalysis. This useful set of data points are called features. The addition of a message to a cover image does not affect the visual appearance of the image but may affect some statistics. The features required for the task of steganalysis should be able to catch these minor statistical disorders that are created during the data hiding process.

In our approach, we extract the first three normalized moments of the characteristic function of the distribution of wavelet coefficients as features. These features are then given to a Fisher Linear Discriminant classifier for classification.

For extraction of features in the Discrete Wavelet Transform domain, we chose three scale decomposition as proposed by Wang and Moulin [12]. Figure 2 shows the levels and selection of subbands for this decomposition.

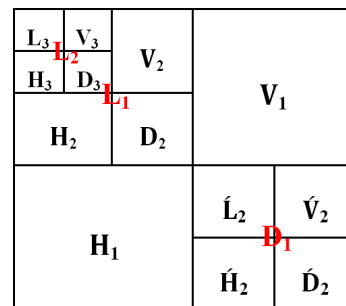
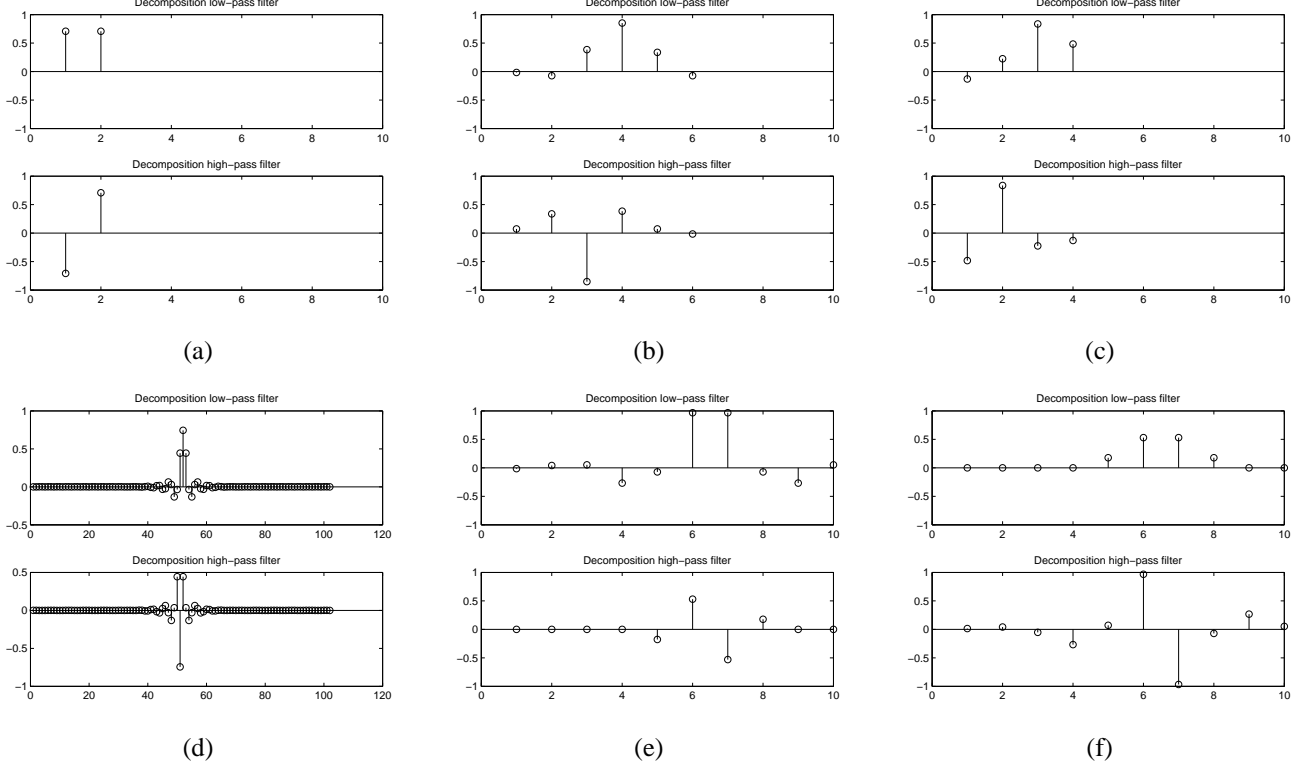


Fig. 2. A three scale wavelet decomposition



**Fig. 3.** Various wavelet decomposition filters used for analysis (a) Daubachies ‘haar’ (b) Coiflets ‘coif1’ (c) Symlets ‘sym2’ (d) Discrete Meyer ‘dmey’ (e) Biorthogonal ‘bior3.5’ (f) Reverse biorthogonal ‘rbio3.5’.

We analyzed six different wavelet filters. The decomposition low-pass and high-pass filters are shown in Figure 3. We obtained nine detail subbands (Horizontal  $\mathbf{H}_i$ , Vertical  $\mathbf{V}_i$  and Diagonal  $\mathbf{D}_i$ ,  $i = 1, 2, 3$ ) and three approximation subbands (Lowpass  $\mathbf{L}_i$ ,  $i = 1, 2, 3$ ). We further decomposed the first scale diagonal subband  $\mathbf{D}_1$  to improve the performance of the features as proposed in [12]. As  $\mathbf{D}_1$  is the finest detail subband and each of its coefficients involves diagonal differences in a four pixel block. So,  $\mathbf{H}_1$ ,  $\mathbf{V}_1$  and  $\mathbf{D}_1$  will contain more information about the difference of differences between neighboring pixels.

Statistical measures are used in our analysis, particularly, the first three normalized moments of the characteristic function are computed. The K-point discrete Characteristic Function (CF) is defined as

$$\Phi(k) = \sum_{m=0}^{M-1} h(m) e^{j \frac{2\pi m k}{K}} \quad (1)$$

where  $\{h(m)\}_{m=0}^{M-1}$  is the  $M$  bin histogram which is an estimate of the PDF,  $p(x)$  of the wavelet coefficients distribution. The  $n^{th}$  absolute moment of discrete CF is defined as

$$M_n^A = \sum_{k=0}^{K/2-1} \Phi(k) \sin^n \left( \frac{\pi k}{K} \right) \quad (2)$$

Finally, the normalized CF moment is defined as

$$\hat{M}_n^A = \frac{M_n^A}{M_0^A} \quad (3)$$

where  $M_0^A$  is the zeroth order moment. We calculated the first three normalized CF moments for each of the 16 subbands, giving a **48-D** feature vector.

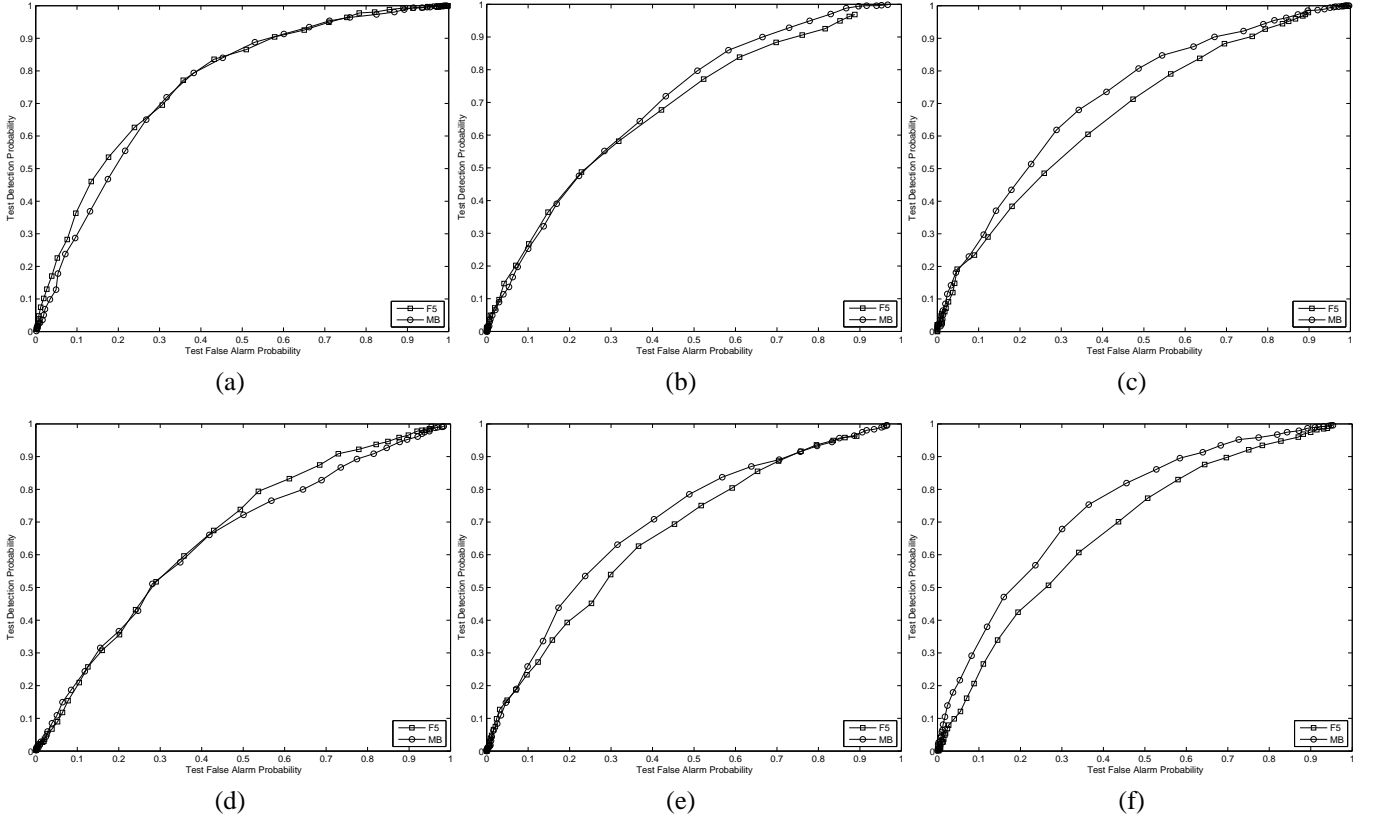
### 3.2. Classifier

We used the two class Fisher Linear Discriminant (FLD) classifier [13]. Let  $\mathbf{x}_i$ ,  $i = 1, \dots, N_x$  and  $\mathbf{y}_j$ ,  $j = 1, \dots, N_y$  represent the samples from each of the two classes of the training set. The within class means are given by

$$\begin{aligned} \mathbf{m}_x &= \frac{1}{N_x} \sum_{i=1}^{N_x} \mathbf{x}_i \\ \mathbf{m}_y &= \frac{1}{N_y} \sum_{j=1}^{N_y} \mathbf{y}_j \end{aligned} \quad (4)$$

The between class mean is

$$\mathbf{m} = \frac{1}{N_x + N_y} \left( \sum_{i=1}^{N_x} \mathbf{x}_i + \sum_{j=1}^{N_y} \mathbf{y}_j \right) \quad (5)$$



**Fig. 4.** The Receiver Operating Characteristics (ROC) curves for the detection of F5 and Model Based steganography algorithms using (a) Daubachies ‘haar’ (b) Coiflets ‘coif1’ (c) Symlets ‘sym2’ (d) Discrete Meyer ‘dmey’ (e) Biorthogonal ‘bior3.5’ (f) Reverse biorthogonal ‘rbio3.5’.

The within class scatter matrix is

$$S_w = M_x M_x^T + M_y M_y^T \quad (6)$$

where  $M_x = \mathbf{x}_i - \mathbf{m}_x$ ,  $M_y = \mathbf{y}_j - \mathbf{m}_y$  are the matrices containing the zero-meaned  $i^{th}$  and  $j^{th}$  samples respectively. The between class scatter matrix is

$$S_b = N_x(\mathbf{m}_x - \mathbf{m})(\mathbf{m}_x - \mathbf{m})^T + N_y(\mathbf{m}_y - \mathbf{m})(\mathbf{m}_y - \mathbf{m})^T. \quad (7)$$

The maximal generalized eigenvalue eigenvector  $\mathbf{e}$  is related to  $S_b$  and  $S_w$  by

$$S_b \mathbf{e} = \lambda S_w \mathbf{e} \quad (8)$$

By projecting the training samples  $\mathbf{x}_i$  and  $\mathbf{y}_j$  onto one dimensional linear subspace  $\mathbf{e}$  ( $x_p = \mathbf{x}_i^T \mathbf{e}$ ,  $y_p = \mathbf{y}_j^T \mathbf{e}$ ), the within class scatter is minimized and the between class scatter is maximized. In any classification problem, this effect is highly desirable as it maintains the discriminability while simultaneously reduces the dimensions of data. An unknown sample  $\mathbf{z}$  can now be tested for its class by projecting it onto the same subspace ( $z_p = \mathbf{z}^T \mathbf{e}$ ) and its class determined on the basis of a threshold  $T_h$ .

## 4. EXPERIMENTAL RESULTS

### 4.1. Image Datasets

#### 4.1.1. Cover Image Dataset

For our experiments, we used 1338 colour images of size 512x384 obtained from the Uncompressed Colour Image Database (UCID) constructed by Schaefer and Stich [14], available at [15]. These images contain a wide range of indoor/outdoor, daylight/night scenes, providing a real and challenging environment for a steganalysis problem. All images were converted to JPEG at 80% quality for our experiments.

#### 4.1.2. F5 Stego Image Dataset

Our first stego image dataset is generated by the steganography software F5 [16], proposed by Andreas Westfeld. F5 steganography algorithm embeds information bits by incrementing and decrementing the values of quantized DCT coefficients from compressed JPEG images [17]. F5 algorithm first compresses the input image with a user defined quality

**Table 1.** Classification results (AUC) for all wavelet filters using FLD for the F5 and Model Based steganography algorithms

Wavelet Filter	F5	MB
Daubachies 'haar'	0.766	0.749
Coiflets 'coif1'	0.666	0.699
Symlets 'sym2'	0.643	0.715
Discrete Meyer 'dmey'	0.653	0.641
Biorthogonal 'bior3.5'	0.654	0.701
Reverse Biorthogonal 'rbio3.5'	0.663	0.736

factor before embedding the message. We chose a quality factor of 80 for stego images. Messages were fully embedded in all images. We chose F5 because recent results in [7], [8], [12] have shown that F5 is harder to detect than other commercially available steganography algorithms.

#### 4.1.3. MB Stego Image Dataset

Our second stego image dataset is generated by the Model Based steganography method [18], proposed by Phil Sallee [19]. The algorithm first breaks down the quantized DCT coefficients of a JPEG image into two parts and then replaces the perceptually insignificant component with the coded message signal. Unlike F5, the Model Based steganography algorithm does not recompress the cover image before embedding. Messages were fully embedded in all images. The model based steganography algorithm has also shown high resistance against steganalysis techniques in [9], [11].

#### 4.2. Evaluation of Results

The Fisher Linear Discriminant classifier [13] was utilized for our experiments. Each steganographic algorithm was analyzed separately for the evaluation of the steganalytic classifier. For a fixed relative message length, we created a database of training images comprising 669 cover and 669 stego images. The wavelet based features were extracted from the training set using various filters according to the procedure explained in Section 3.1. The FLD classifier was then tested on the features extracted from a different database of test images comprising 669 cover and 669 stego images. The Receiver Operating Characteristics (ROC) curves, which give the variation of the Detection Probability ( $P_d$ , the fraction of correctly classified stego images) with the False Alarm Probability ( $P_f$ , the fraction of stego images wrongly classified as cover image), were computed for each steganographic algorithm and embedding rate. The area under the ROC curve (AUC) was measured to determine the overall classification accuracy. Figures 4(a)-(f) show the obtained ROC curves for both F5 and Model based steganography algorithms.

Table 1 summarizes the classification results. We observe that the 'haar' filter outperforms all filters for both F5 and

Model Based steganography algorithms. This shows that the 'haar' filter is most sensitive to the variations in an image due to embedding of a message. Another observation is that the 'haar' and 'dmey' filters are more sensitive to the embedding variations due to F5 steganography algorithms. Similarly, the 'coif1', 'sym2', 'bior3.5' and 'rbio3.5' are more sensitive to the embedding variations due to Model Based steganography algorithm.

### 5. CONCLUSION

This paper presents an evaluation of wavelet filters performance for the task of steganalysis. Six different wavelet filters from the Daubachies, Coiflets, Symlets, Discrete Meyer, Biorthogonal and Reverse Biorthogonal families were analyzed. The experiments revealed that the simplest wavelet filter 'haar' gives the best performance for wavelet based steganalysis for both commercially available F5 and Model Based steganography algorithms. Further research may be done in the exploration of custom made wavelet filters specific for the task of steganalysis or using a combination of low-pass and high-pass filters of two different types.

### 6. REFERENCES

- [1] N.F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *IEEE Computer*, vol. 31, no. 2, pp. 26–34, 1998.
- [2] G. J. Simmons, "Prisoners' problem and the subliminal channel," in *CRYPTO83-Advances in Cryptology*, 1984, pp. 51–67.
- [3] J. Fridrich, M. Goljan, and D. Hoge, "Steganalysis of jpeg images: Breaking the f5 algorithm," in *Proc. of 5th International Workshop on Information Hiding*, Noordwijkerhout, Netherlands, October 2002, pp. 310–323.
- [4] H. A. Aboalsamh, S. A. Dokheekh, H. I. Mathkour, and G. M. Assassa, "Breaking the f5 algorithm: An improved approach," in *Egyptian Computer Science Journal*, January 2007, vol. 29, pp. 1–9.
- [5] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *3rd Information Hiding Workshop*, Germany, 1999, pp. 61–76.
- [6] J. Fridrich, M. Goljan, and D. Hoge, "Attacking the outguess," in *Proc. ACM Workshop on Multimedia and Security*, Juan-les-Pins, France, December 2002.
- [7] I. Avcibas, N. Memon, and B. Sankur, "Image steganalysis with binary similarity measures," in *IEEE International Conference on Image Processing*, Rochester, New York, September 2002.

- [8] H. Farid, "Detecting hidden messages using higher-order statistical models," in *Proc. of the IEEE International Conference on Image Processing*, 2002, vol. 2, pp. 905–908.
- [9] J. Fridrich, "Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes," in *I. S. Moskowitz (Ed.): Information Hiding*. 2004, pp. 67–81, LNCS 2137, Springer-Verlag, Berlin Heidelberg.
- [10] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," in *IEEE Transactions on Image Processing*, 2003, vol. 12, pp. 221–229.
- [11] M. Kharrazi, T.H. Sencar, and N. Memon, "Benchmarking steganographic and steganalysis techniques," in *Proc. of SPIE Electronic Imaging, Security, Steganography and Watermarking of Multimedia Contents VII*, 2005.
- [12] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," in *IEEE Transactions on Information Forensics and Security*, 2007, vol. 2.
- [13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2nd edition, 2001.
- [14] G. Schaefer and M. Stich, "Ucid - an uncompressed colour image database," in *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, San Jose, USA, 2004, vol. 2, pp. 472–480.
- [15] UCID - Uncompressed Colour Image Database. [Online]. Available, ," <http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html>, visited on 02/08/08.
- [16] Steganography Software F5. [Online]. Available, ," <http://wwrn.inf.tu-dresden.de/~westfeld/f5.html>, visited on 02/08/08.
- [17] A. Westfeld, "F5 – a steganographic algorithm: High capacity despite better steganalysis," in *I.S. Moskowitz, (Ed.): Information Hiding. 4th International Workshop*. April 2001, pp. 289–302, LNCS, Springer-Verlag, Berlin Heidelberg.
- [18] Model Based JPEG Steganography Demo. [Online]. Available, ," <http://www.philsaltee.com/mbsteg/index.html>, visited on 02/08/08.
- [19] P. Sallee, "Model based steganography," in *International Workshop on Digital Watermarking*, Seoul, Korea, October 2003, pp. 174–188.