A New Hybrid DCT and Contourlet Transform Based JPEG Image Steganalysis Technique

Zohaib Khan and Atif Bin Mansoor

College of Aeronautical Engineering,
National University of Sciences & Technology, Pakistan
zohaibkh_27@yahoo.com, atif-cae@nust.edu.pk

Abstract. In this paper, a universal steganalysis scheme for JPEG images based upon hybrid transform features is presented. We first analyzed two different transform domains (Discrete Cosine Transform and Discrete Contourlet Transform) separately, to extract features for steganalysis. Then a combination of these two feature sets is constructed and employed for steganalysis. A Fisher Linear Discriminant classifier is trained on features from both clean and steganographic images using all three feature sets and subsequently used for classification. Experiments performed on images embedded with two variants of F5 and Model based steganographic techniques reveal the effectiveness of proposed steganalysis approach by demonstrating improved detection for hybrid features.

Keywords: Steganography, Steganalysis, Information Hiding, Feature Extraction, Classification.

1 Introduction

The word steganography comes from the Greek words *steganos* and *graphia*, which together mean 'hidden writing' [1]. Steganography is being used to hide information in digital images and later transfer them through the internet without any suspicion. This poses a serious threat to both commercial and military organizations as regards to information security. Steganalysis techniques aim at detecting the presence of hidden messages from inconspicuous stego images.

Steganography is an ancient subject, with its roots lying in ancient Greece and China, where it was already in use thousands of years ago. The prisoners' problem [2] well defines the modern formulation of steganography. Two accomplices Alice and Bob are in a jail. They wish to communicate in order to plan to break the prison. But all communication between the two is being monitored by the warden, Wendy, who will put them in a high security prison if they are suspected of escaping. Specifically, in terms of a steganography model, Alice wishes to send a secret message $\boldsymbol{m}$ to Bob. For this, she hides the secret message $\boldsymbol{m}$ using a shared secret key $\boldsymbol{k}$ into a cover-object $\boldsymbol{c}$ to obtain the stego-object $\boldsymbol{s}$. The stego-object $\boldsymbol{s}$ is then sent by Alice through the public channel to Bob, $\boldsymbol{m}$ unnoticed by Wendy. Once Bob receives the stego-object $\boldsymbol{s}$, he is able to recover the secret message $\boldsymbol{m}$ using the shared secret key $\boldsymbol{k}$.

A.-B. Salberg, J.Y. Hardeberg, and R. Jenssen (Eds.): SCIA 2009, LNCS 5575, pp. 321–330, 2009.

Steganography and cryptography are closely related information hiding techniques. The purpose of cryptography is to scramble a message so that it cannot be understood, while that of steganography is to hide a message so that it cannot be seen. Generally, a message created with cryptographic tools will raise the alarm on a neutral observer while a message created with steganographic tools will not. Sometimes, steganography and cryptography are combined in a way that the message may be encrypted before hiding to provide additional security.

Steganographers who intend to hide communications are countered by steganalysts who intend to reveal it. The specific field to counter steganography is known as *steganalysis*. The 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 further identification of the steganography tool to extract the secret message from the stego file comes under the field of cryptanalysis.

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 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 [4], [5], [6] and [7]. 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 [8], [9], [10] and [11]. Our research work is concentrated on the second approach due to its wide applicability.

In this paper, we propose a steganalysis technique by extracting features from two transform domains; the discrete contourlet transform and the discrete cosine transform. These features are investigated individually and combinatorially. 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. [3].

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 [4]. This method was further improved by Aboalsamh et al. [5] by determining the optimal value of the message length estimation parameter β. 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 χ^2 statistic [6]. A steganalysis scheme specific to the embedding algorithm Outguess is proposed in [7], by making use of the assumption that the embedding of a message in a stego image will be different than 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 [8]. Farid suggested that embedding of a message alters the higher order statistics calculated from a multi-scale wavelet decomposition [9]. 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 [10]. 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 [11].

3 Proposed Approach

3.1 Feature Extraction

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 first extract features in the discrete contourlet transform domain, followed by the discrete cosine transform domain and finally combine both extracted features to make a hybrid feature set.

Discrete Contourlet Tranform Features. The contourlet transform is a new two-dimensional extension of the wavelet transform using multiscale and directional filter banks [13]. For extraction of features in the Discrete Contourlet Transform domain, we decomposed image into three pyramidal levels and 2^n directions where $n = 0, 2, 4$. Figure 1 shows the levels and selection of subbands for this decomposition. For the laplacian pyramidal decomposition stage, the *'Haar'* filter was used. For the directional decomposition stage the *'PKVA'* filter was used. In each scale from coarse to fine, the number of directions are 1,4,and 16. By applying the pyramidal directional filter bank decomposition and ignoring the finest lowpass approximation subband, we obtained a total of 23 subbands.

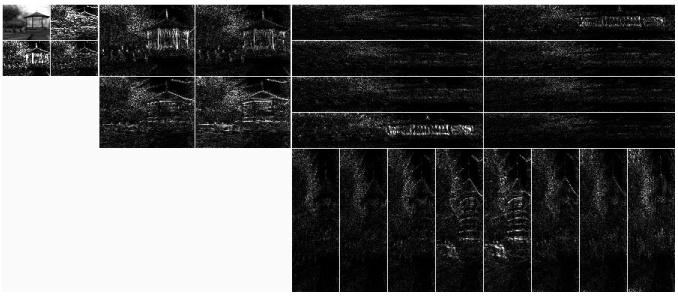


Fig. 1. A three level contourlet decomposition

Various 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

$$\Phi(k) = \sum_{m=0}^{M-1} h(m) e^{\{\frac{j2\pi mk}{K}\}} \ . \tag{1}$$

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

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

Finally, the normalized CF moment is defined as

$$\hat{M}_n^A = \frac{M_n^A}{M_0^A} \ . \tag{3}$$

where M_0^A is the zeroth order moment. We calculated the first three normalized CF moments for each of the 23 subbands, giving a **69-D** feature vector.

DCT Based Features. The DCT based feature set is constructed following the approach of Fridrich [10]. A vector functional $\boldsymbol{F}$ is applied to the JPEG image J_1. This image is then decompressed to the spatial domain, cropped by 4 pixels in each direction and recompressed with the same quantization table as J_1 to obtain J_2. The vector functional $\boldsymbol{F}$ is then applied to J_2. The final feature f is obtained as the L_1 norm of the difference of the functional applied to J_1 and J_2.

$$f = \|\boldsymbol{F}(J_1) - \boldsymbol{F}(J_2)\|_{L_1} \ . \tag{4}$$

The rational behind this procedure is that the recompression after cropping by 4 pixels does not see the previous JPEG compression's 8×8 block boundary and thus it is not affected by the previous quantization and hence embedding in the DCT domain. So, J_2 can be thought of as an approximation to its cover image.

We calculated the global, individual and dual histograms of the DCT coefficient array $d_{(k)}(i,j)$ as the first order functionals. The symbol $d_{(k)}(i,j)$ denotes the $(i,j)^{th}$ quantized DCT coefficient $(i,j=1,2,...,8)$ in the *k-th* block, $(k=1,2,...,B)$. The global histogram of all 64B DCT coefficients is given as, $H(m)_{m=L}^{R}$, where $L=min_{k,i,j}d_{(k)}(i,j)$ and $R=max_{k,i,j}d_{(k)}(i,j)$. We computed $H/\left\|H\right\|_{L_1}$, the normalized global histogram of DCT coefficients as the first functional.

Steganographic techniques that preserve global DCT coefficients histogram may not necessarily preserve the histogram of individual DCT modes. So, we calculated $h^{ij}/\left\|h^{ij}\right\|_{L_1}$, the normalized individual histograms $h(m)_{m=L}^{R}$ of 5 low frequency DCT modes, $(i,j)=(2,1)$, $(3,1)$, $(1,2)$, $(2,2)$, $(1,3)$ as the next five functionals.

The dual histogram is an 8×8 matrix which indicates the number of how many times the value *'d'* occurs as the $(i,j)^{th}$ DCT coefficient over all blocks B in the image. We computed $g_{ij}^{d}/\left\|g_{ij}^{d}\right\|_{L_1}$, the normalized dual histograms where $g_{ij}^{d}=\sum_{k=1}^{B}\delta(d,d_{(k)}(i,j))$ for 11 values of $d=-5,-4,...,4,5$.

Inter block dependency is captured by the second order features *variation* and *blockiness*. Most steganographic techniques add entropy to the DCT coefficients which is captured by the *variation* (V)

$$V=\frac{\sum_{i,j=1}^{8}\sum_{k=1}^{|I_r|-1}|d_{I_r(k)}(i,j)-d_{I_r(k+1)}(i,j)|+\sum_{i,j=1}^{8}\sum_{k=1}^{|I_c|-1}|d_{I_c(k)}(i,j)-d_{I_c(k+1)}(i,j)|}{|Ir|+|Ic|}. \tag{5}$$

where I_r and I_c denote the vectors of block indices while scanning the image 'by rows' and 'by columns' respectively.

Blockiness is calculated from the decompressed JPEG image and is a measure of discontinuity along the block boundaries over all DCT modes over the whole image. The L_1 and L_2 *blockiness* $(B_\alpha,\alpha=1,2)$ is defined as

$$B_\alpha=\frac{\sum_{i=1}^{\lfloor(M-1)/8\rfloor}\sum_{j=1}^{N}|x_{8i,j}-x_{8i+1,j}|^{\alpha}+\sum_{j=1}^{\lfloor(N-1)/8\rfloor}\sum_{i=1}^{M}|x_{i,8j}-x_{i,8j+1}|^{\alpha}}{N\lfloor(M-1)/8\rfloor+M\lfloor(N-1)/8\rfloor} \tag{6}$$

where $x_{i,j}$ are the grayscale intensity values of an image with dimensions $M\times N$.

The final DCT based feature vector is **20-D** (Histograms: 1 global, 5 individual, 11 dual. *Variation*: 1. *Blockiness*: 2).

Hybrid Features. After extracting the features in the discrete cosine transform and the discrete contourlet transform domain, we finally combine the extracted feature sets into one hybrid feature set, giving a **89-D** feature vector, (69 CNT + 20 DCT).

4 Experimental Results

4.1 Image Datasets

Cover Image Dataset. For our experiments, we used 1338 grayscale 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.

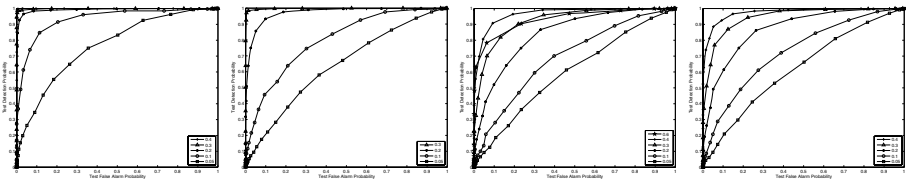
F5 Stego Image Dataset. Our first stego image dataset is generated by the steganography software F5 [16], proposed by Andreas Westfeld. F5 steganography algorithm embeds information bits by incrementing and decrementing the values of quantized DCT coefficients from compressed JPEG images [17]. F5 also uses an operation known as 'matrix embedding' in which it minimizes the amount of changes made to the DCT coefficients necessary to embed a message of certain length. Matrix embedding has three parameters (c, n, k), where c is the number of changes per group of n coefficients, and k is the number of embedded bits. These parameter values are determined by the embedding algorithm.

F5 algorithm first compresses the input image with a user defined quality factor before embedding the message. We chose a quality factor of 80 for stego images. Messages were successfully embedded at rates of 0.05, 0.10, 0.20, 0.3, 0.40 and 0.60 bpc (bits per non-zero DCT coefficients). We chose F5 because recent results in [8], [9], [12] have shown that F5 is harder to detect than other commercially available steganography algorithms.

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

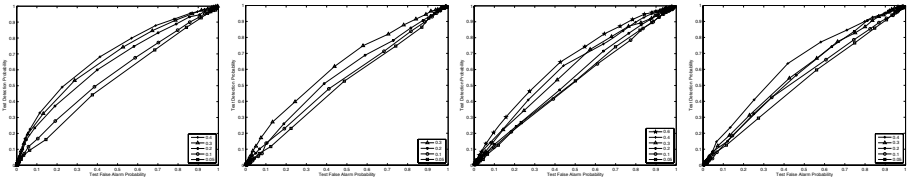
Table 1. The number of images in the stego image datasets given the message length. F5 with matrix embedding turned off $(1, 1, 1)$ and turned on (c, n, k). Model based steganography without deblocking (MB1) and with deblocking (MB2). (U = unachievable rate).

Embedding Rate (bpc)	**F5** $(1, 1, 1)$	**F5** (c, n, k)	**MB1**	**MB2**
0.05	1338	1338	1338	1338
0.10	1338	1338	1338	1338
0.20	1338	1337	1338	1334
0.30	1337	1295	1338	1320
0.40	1332	5	1338	1119
0.60	5	U	1332	117
0.80	U	U	60	U



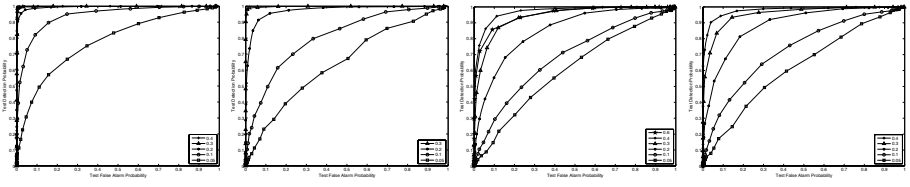
(a) (b) (c) (d)

Fig. 2. ROC curves using DCT based features. (a) F5 (without matrix embedding) (b) F5 (with matrix embedding) (c) MB1 (without deblocking) (d) MB2 (with deblocking).



(a) (b) (c) (d)

Fig. 3. ROC curves using CNT based features. (a) F5 (without matrix embedding) (b) F5 (with matrix embedding) (c) MB1 (without deblocking) (d) MB2 (with deblocking).



(a) (b) (c) (d)

Fig. 4. ROC curves using Hybrid features. (a) F5 (without matrix embedding) (b) F5 (with matrix embedding) (c) MB1 (without deblocking) (d) MB2 (with deblocking).

with the coded message signal. The algorithm has two types; MB1 is normal steganography and MB2 is steganography with deblocking. The deblocking algorithm adjusts the unused coefficients to reduce the blockiness of the resulting image to the original blockiness. Unlike F5, the Model Based steganography algorithm does not recompress the cover image before embedding. We embed at rates of 0.05, 0.10, 0.20, 0.3, 0.40 0.60 and 0.80 bpc. The model based steganography algorithm has also shown high resistance against steganalysis techniques in [3], [10].

The reason for choosing the message length proportional to the number of non-zero DCT coefficients was to create a stego image database for which the steganalysis is roughly of the same level of difficulty. We further carried out embedding at different rates to observe the steganalysis performance for messages of varying length. It can be seen in Table 1 that the Model based steganography is more efficient in embedding as compared to F5; since longer messages can be accommodated in images using Model based steganography.

Table 2. Classification results (AUC) using FLD for all embedding rates. F5 with matrix embedding turned off $(1, 1, 1)$ and turned on (c, n, k). Model based steganography without deblocking (MB1) and with deblocking (MB2). (U = unachievable rate).

Rate (bpc)	**F5** $(1, 1, 1)$	**F5** (c, n, k)	**MB1**	**MB2**	
0.05	0.769	0.643	0.611	0.591	DCT
0.05	0.555	0.511	0.529	0.518	CNT
0.05	0.789	0.632	0.624	0.585	HYB
0.10	0.924	0.795	0.721	0.686	DCT
0.10	0.589	0.543	0.511	0.508	CNT
0.10	0.936	0.800	0.723	0.681	HYB
0.20	0.989	0.968	0.860	0.829	DCT
0.20	0.639	0.572	0.570	0.541	CNT
0.20	0.990	0.971	0.886	0.851	HYB
0.30	0.998	0.997	0.934	0.914	DCT
0.30	0.688	0.629	0.590	0.576	CNT
0.30	0.996	0.996	0.953	0.935	HYB
0.40	1.000	U	0.963	0.962	DCT
0.40	0.697	U	0.617	0.619	CNT
0.40	0.997	U	0.978	0.974	HYB
0.60	U	U	0.984	U	DCT
0.60	U	U	0.667	U	CNT
0.60	U	U	0.990	U	HYB

4.2 Evaluation of Results

The Fisher Linear Discriminant classifier [20] 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. Both DWT based features (DWT) and DCT based features (DCT) were extracted from the training set and were combined to form a Joint feature set (JNT), 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 2-4 give the obtained ROC curves for the steganographic techniques under test for different embedding rates. Note that due to the space limitation, these figures are displayed in small size. However, readers are encouraged to take a look by using zoom to 400%. We observe that the DCT based features outperform the CNT based features for all embedding rates. As could be expected, the

detection of F5 without matrix embedding is better than F5 with matrix embedding since the matrix embedding operation significantly reduces detectability at the expense of message capacity.

Table 2 summarizes the classification results. For F5 without matrix embedding, the proposed Hybrid transform features dominate both DCT and CNT based features for embedding rates till 0.20 bpc. For higher embedding rates the DCT based features perform better. For F5 with matrix embedding, both the proposed hybrid features and the DCT based features are close competitors, though the former performs better at some embedding rates.

For MB1 algorithm (without deblocking), the proposed hybrid features outperform both the DCT and CNT based features for all embedding rates. For MB2 algorithm (with deblocking), the hybrid features perform better compared to both CNT and DCT based features for embedding rates greater than 0.10 bpc. It is observed that the detection of MB1 is better than MB2, as the deblocking algorithm in MB2 reduces the blockiness of the stego image to match the original image.

5 Conclusion

This paper presents a new DCT and CNT based hybrid features approach for universal steganalysis. DCT and CNT based statistical features are investigated individually, followed by research on combined features. The Fisher Linear Discriminant classifier is employed for classification. The experiments were performed on image datasets with different embedding rates for F5 and Model based steganography algorithms. Experiments revealed that for JPEG images the DCT is a better choice for extraction of features as compared to the CNT. The experiments with hybrid transform features reveal that the extraction of features in more than one transform domain improves the steganalysis performance.

References

1. McBride, B.T., Peterson, G.L., Gustafson, S.C.: A new Blind Method for Detecting Novel Steganography. Digital Investigation 2, 50–70 (2005)
2. Simmons, G.J.: 'Prisoners' Problem and the Subliminal Channel. In: CRYPTO 1983-Advances in Cryptology, pp. 51–67 (1984)
3. Kharrazi, M., Sencar, T.H., Memon, N.: Benchmarking Steganographic and Steganalysis Techniques. In: Proc. of SPIE Electronic Imaging, Security, Steganography and Watermarking of Multimedia Contents VII, San Jose, California, USA (2005)
4. Fridrich, J., Goljan, M., Hogea, D.: Steganalysis of JPEG images: Breaking the F5 Algorithm. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 310–323. Springer, Heidelberg (2003)
5. Aboalsamh, H.A., Dokheekh, S.A., Mathkour, H.I., Assassa, G.M.: Breaking the F5 Algorithm: An Improved Approach. Egyptian Computer Science Journal 29(1), 1–9 (2007)

6. Westfeld, A., Pfitzmann, A.: Attacks on Steganographic Systems. In: Proc. 3rd Information Hiding Workshop, Dresden, Germany, pp. 61–76 (1999)
7. Fridrich, J., Goljan, M., Hogea, D.: Attacking the OutGuess. In: Proc. ACM Workshop on Multimedia and Security 2002. ACM Press, Juan-les-Pins (2002)
8. Avcibas, I., Memon, N., Sankur, B.: Image Steganalysis with Binary Similarity Measures. In: Proc. of the IEEE International Conference on Image Processing, Rochester, New York (September 2002)
9. Farid, H.: Detecting Hidden Messages Using Higher-order Statistical Models. In: Proc. of the IEEE International Conference on Image Processing, vol. 2, pp. 905–908 (2002)
10. Fridrich, J.: Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes. In: Moskowitz, I.S. (ed.) Information Hiding 2004. LNCS, vol. 2137, pp. 67–81. Springer, Heidelberg (2005)
11. Avcibas, I., Memon, N., Sankur, B.: Steganalysis Using Image Quality Metrics. IEEE Transactions on Image Processing 12(2), 221–229 (2003)
12. Wang, Y., Moulin, P.: Optimized Feature Extraction for Learning-Based Image Steganalysis. IEEE Transactions on Information Forensics and Security 2(1) (2007)
13. Po, D.-Y., Do, M.N.: Directional Multiscale Modeling of Images Using the Contourlet Transform. IEEE Transactions on Image Processing 15(6), 1610–1620 (2006)
14. Schaefer, G., Stich, M.: UCID - An Uncompressed Colour Image Database. In: Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia, San Jose, USA, pp. 472–480 (2004)
15. UCID – Uncompressed Colour Image Database, `http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html` (visited on 02/08/08)
16. Steganography Software F5, `http://wwwrn.inf.tu-dresden.de/~westfeld/f5.html` (visited on 02/08/08)
17. Westfeld, A.: F5 – A Steganographic Algorithm: High capacity despite better steganalysis. In: Moskowitz, I.S. (ed.) IH 2001. LNCS, vol. 2137, pp. 289–302. Springer, Heidelberg (2001)
18. Model Based JPEG Steganography Demo, `http://www.philsallee.com/mbsteg/index.html` (visited on 02/08/08)
19. Sallee, P.: Model-based steganography. In: Kalker, T., Cox, I., Ro, Y.M. (eds.) IWDW 2003. LNCS, vol. 2939, pp. 154–167. Springer, Heidelberg (2004)
20. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons, New York (2001)