The Influence of the Image Basis on Modeling and Steganalysis Performance

Valentin Schwamberger[1], Pham Hai Dang Le[2], Bernhard Schölkopf[1], and Matthias O. Franz[2]

[1] Max Planck Institute for Biological Cybernetics, Spemannstr. 38,
72076 Tübingen, Germany,
vschwamb@gmail.com, bs@tuebingen.mpg.de,
[2] HTWG Konstanz, Institute for Optical Systems, Brauneggerstr. 55,
78462 Konstanz, Germany,
{mfranz,dangle}@htwg-konstanz.de

Abstract. We compare two image bases with respect to their capabilities for image modeling and steganalysis. The first basis consists of wavelets, the second is a Laplacian pyramid. Both bases are used to decompose the image into subbands where the local dependency structure is modeled with a linear Bayesian estimator. Similar to existing approaches, the image model is used to predict coefficient values from their neighborhoods, and the final classification step uses statistical descriptors of the residual. Our findings are counter-intuitive on first sight: Although Laplacian pyramids have better image modeling capabilities than wavelets, steganalysis based on wavelets is much more successful. We present a number of experiments that suggest possible explanations for this result.

1 Introduction

Most steganalytic methods are not capable of detecting general steganographic manipulations in images (universal steganalysis), since they are tuned to specific steganographic algorithms. The few currently available universal steganalytic algorithms [13, 8, 11, 2] are relatively insensitive towards small embeddings. This is due to the problem of detecting a tiny manipulation (the embedded data) in a large amplitude signal (the carrier image).

The large amplitude of the carrier signal can be largely reduced by applying an image model to a suitably transformed image. The image model is capable of predicting transform coefficients from their local neighborhoods [4] based on the coefficient statistics of the image. Since an embedded message cannot be predicted from the neighborhood statistics of the image, it must be part of the prediction error of the model [13]. Thus, by analyzing the prediction error instead of the whole image, we effectively remove most of the carrier signal. The residual is much more affected by the embedding manipulation than the full image which results in a better detectability.

The initial image transform determines the basis in which the image is modeled and in which the residual is characterized by a suitable set of statistical descriptors which constitute the input to a final classifier stage. In such a steganalyzer architecture, a plausible hypothesis can be stated as follows: “The best image basis (or the best associated subband transform) is that which leads to the image model with the highest predictability since this most effectively removes the carrier from a potential stego image”. Here, we show that this is not the case, and provide some hints on the possible reasons for this counter-intuitive result.

Our study is based on a modified version of the well-known steganalyzer of Lyu and Farid [13] which we describe in the next section. The investigated image bases are QMF wavelets [18] and Laplacian pyramids [1]. In Sect. 3, we present our results on image modeling and steganalysis performance. Additional experiments for explaining these results are discussed in Sect. 4. We conclude with a brief summary in Sect. 5.

2 Lyu and Farid’s Algorithm and Modifications

The input of Lyu and Farid’s algorithm is an image in its pixel representation. Originally, a wavelet pyramid is built for each color channel (as shown in the upper path in Fig. 1). Alternatively, the image can be decomposed into a Laplacian pyramid described later (lower path in Fig. 1). Quadrature mirror filters are used for building the wavelet pyramid [18] with a quadrature mirror filter of width 9. We get $3(3s + 1)$ subbands for an RGB image and a pyramid with s scales and three orientation subbands, i. e. diagonal, vertical, and horizontal orientation. In the case of the alternative Laplacian pyramid representation [1], we use a standard binomial filter of width 5 to obtain the lowband approximation of the image. The number of pyramid levels was chosen to be the same as in the wavelet pyramid, but—since the Laplacian pyramid does not decompose the image according to orientation—we have only one subband per pyramid level which results in overall $3s$ subbands for color images.

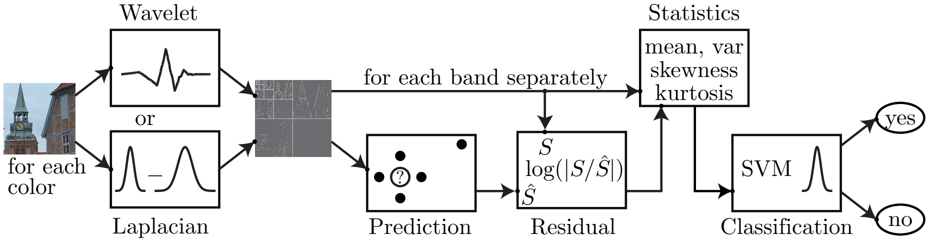
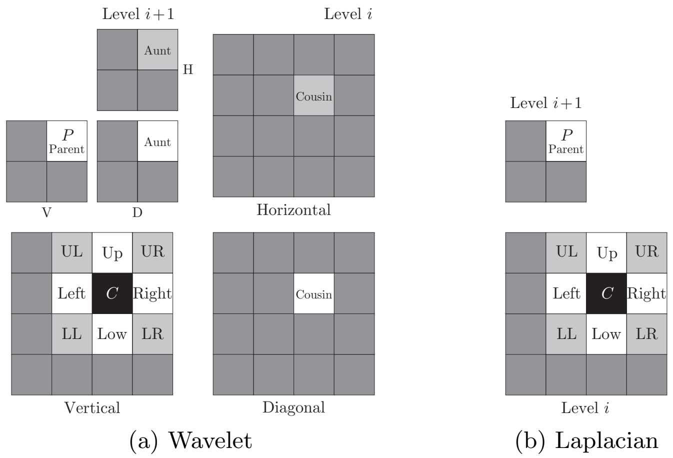


Fig. 1. The algorithm—schematics



(a) Wavelet (b) Laplacian

Fig. 2. Neighborhood structure for image modeling (color neighbors not included). The central coefficient to be predicted is C, the light gray neighbors can be optionally included but did not lead to significantly different results.

For predicting coefficients from their neighborhood, we need to specify a neighborhood structure for each image representation which is shown in Fig. 2. The neighborhood structure for wavelets is the same as in [13] (cf. Fig. 2a), the Laplacian neighborhood is constructed analogously (cf. Fig. 2b), but without orientation neighbors. Both representations contain the corresponding central coefficient from the other color channels in their neighborhoods (not shown in Fig. 2). Due to only including the neighboring coefficients from closest orientations on the same scale (hence including horizontal and vertical coefficients for predicting the diagonal subband, but only diagonal coefficients for both the horizontal and vertical subbands), and correspondingly only one (diagonal) or two neighbors (horizontal and vertical) from the coarser scales, neighborhoods in the wavelet representation contain 9 coefficients, in the Laplacian representation 7 coefficients.

The predictions are computed with linear regression applied to each subband separately, i. e., the magnitude of the central coefficient is obtained as a weighted sum of the magnitudes of its neighboring coefficients greater than a given threshold: It has been shown empirically that only the magnitudes of coefficients are correlated, and the correlation decreases for smaller magnitudes [4]. The weight sets over all subbands thus constitute the image model. In their original approach, Lyu & Farid used standard least-squares regression for this purpose. In our implementation, we use Gaussian process (GP) regression [14, 15] instead after normalizing all subband coefficients to the interval $[0, 1]$. This approach leads to slightly more robust, but essentially comparable results for the purpose of this study. GP regression needs an additional model selection step for estimating the noise content in the image. For that, we use Geisser's surrogate predictive probability [7]. It is computed on a subset of of the coefficients: The finest scales are subsampled by a factor of 5 and the coarser by a factor of 3, each in both directions. Details on this regression technique can be found in [15].

Each estimator is trained and used for prediction on the same subband. Thus, training and test set coincide for this application. From the predicted coefficients $\widehat{S}$, small coefficients with amplitude below a threshold of $t = 1/255$ are set to zero. For reconstructing complete images, the algebraic signs are transferred from the original to the predicted subband coefficients. The residual r is computed by taking the logarithm of the coefficients of the input image transform S and the predicted coefficients $\widehat{S}$ and subtracting them subsequently, hence $r = \log S - \log \widehat{S}$.

Next, the four lowest statistical moments—i. e. mean, standard deviation, skewness, and kurtosis—of the subband coefficients (called *marginal statistics* in [13]) and of the subband residuals (called *error statistics*) are computed, again for each color and subband separately. Finally, all these independently normalized statistics serve as feature inputs for a support vector machine [17]. In this study, we use $s = 3$ pyramid levels which results in a 120-dimensional feature vector for the wavelet representation and in a 48-dimensional vector for the Laplacian decomposition. The final classification was done with a 1-norm soft margin non-linear C-SVM using a Gaussian kernel. The choice of the parameter C of the SVM and the width σ of the Gaussian kernel was based on a new paired cross-validation procedure described elsewhere ([16], in preparation). The SVM is tunable in order to adapt the rate of false alarms and the detection rate.

3 Comparison between Wavelet and Laplacian Basis

3.1 Image Modeling Performance

The prediction quality of the image model is measured in terms of the *explained variance* in pixel space

$$\mathcal{V}_{\mathrm{expl}} \equiv \frac{\mathcal{V}_{\mathrm{img}} - \mathcal{V}_{\mathrm{err}}}{\mathcal{V}_{\mathrm{img}}}$$

with the variance $\mathcal{V}_{\mathrm{img}} \equiv (1/n) \sum_{i,j} \big(S(x_i, y_j) - \bar{S} \big)^2$ of the n image pixels $S(x_i, y_j)$ (with mean $\bar{S}$) and the mean square error $\mathcal{V}_{\mathrm{err}} \equiv (1/n) \sum_{i,j} \big(S(x_i, y_j) - \widehat{S}(x_i, y_j) \big)^2$ where $\widehat{S}(x_i, y_j)$ are the predicted pixel values and the (i, j) run over all pixels in the image[3]. In addition, we provide explained variances for each analyzed image scale separately to highlight the relative contribution of each image scale to the overall error. In this case, variances, errors and predictions are computed in transform coefficient space instead of pixel space, and the (i, j) run over all coefficients belonging to a given scale.

[3] We prefer explained variance over the frequently used mean square error since it provides a relative measure of image modeling performance and thus is independent of the actual scaling of the pixel values.

We compared the explained variances of the two image bases on the familiar Brodatz texture database [3] which contains 111 640 × 640-sized grayscale images scanned off black and white prints. In addition, we tested both bases on an image database containing more than 1600 never compressed RGB color images provided by the German Federal Office for Information Security. Although textures are not representative for natural images, they constitute a good testbed for local Markov random field (MRF) type image models such as ours since they are statistically uniform at a limited range of scales and orientations and thus help to reveal potential weaknesses of a model which otherwise could remain invisible in natural images with their variable mixture of local textures. Typically, a higher performance of a local MRF-type model on a texture database leads to a higher performance on natural images which was also the case in our tests.

Fig. 3 shows that the Laplacian image basis outperforms the wavelet basis significantly in terms of explained variance, even if training and test region of the images were not the same. This happened consistently, both in pixel space (first bar group, "pixel space reconstruction") and across the different scales of the Laplace or wavelet decomposition (bar groups numbered 1–4). For RGB images, the advantages of the Laplacian basis are less pronounced but still significant, since the high correlations between the color channels are exploited by the models as well and thus lead to smaller differences in their prediction performance, see Fig. 4. The better prediction performance of the Laplace basis can be attributed to two factors: (1) Laplace coefficients are higher correlated with their neighbourhood than wavelet coefficient magnitudes and thus are easier to predict; (2) The Laplace pyramid is overcomplete by a factor of 4/3 which allows for a more finely grained modeling of the local dependency structure.

3.2 Steganalysis Performance

The wavelet and the Laplacian image models were used for determining both marginal and residual statistics for the above-mentioned image database of never compressed color images. This is known to be the most difficult setting for steganalysis, as the entropies of the images remain high. For instance, JPEG artifacts contained in the images from previous compression simplify steganalysis [11]. Different embedding algorithms and rates were used for creating sets of stego images from these clean images.

The comparison of the distributions of the residuals for a clean color image and its corresponding stego version can be seen in Fig. 5. Here, for the sake of easily recognizable differences, the complete least significant bit plane was replaced by white noise, serving as a representative of a very simple steganogram. In Fig. 6, the corresponding distributions for the same color image are shown for the Laplacian image model. The differences are very small, compared to the distributions computed with the wavelet model.

From these distributions, we computed the statistical moments for every color image that serve as associated feature vectors. The dimensionality of these vectors was 120 for wavelet decomposition, and 48 for Laplacian decomposition. After normalizing the components of these vectors independently, we carefully

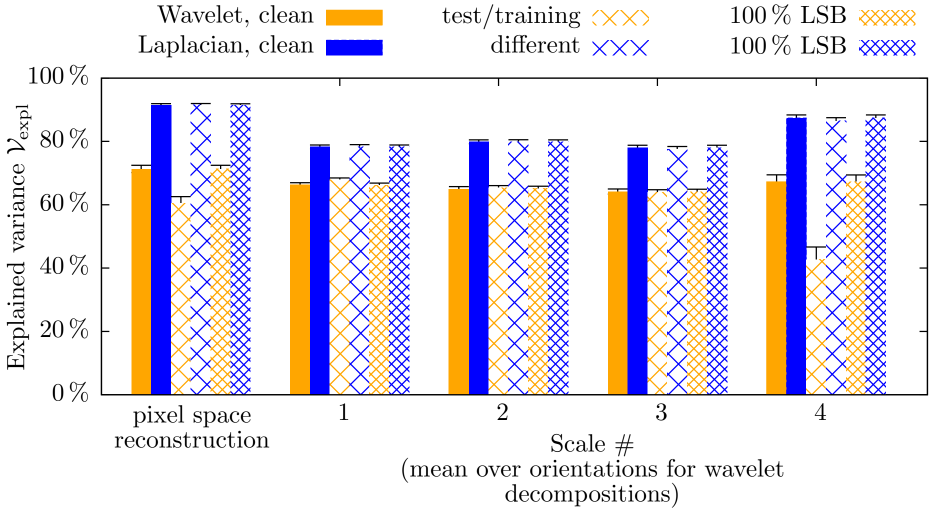


Fig. 3. Explained variances $\mathcal{V}_{\text{expl}}$ for Wavelet and Laplacian decompositions averaged over the Brodatz database, for each subband scale and pixel space reconstruction (leftmost bar group). The black bars indicate the standard error of the mean over the database. Results based on the wavelet representation are shown in orange, results based on the Laplacian representation in blue. The title "test/training different" in the legend belongs to both representations and indicates that training and prediction of the image model were carried out on different sections of the same image.

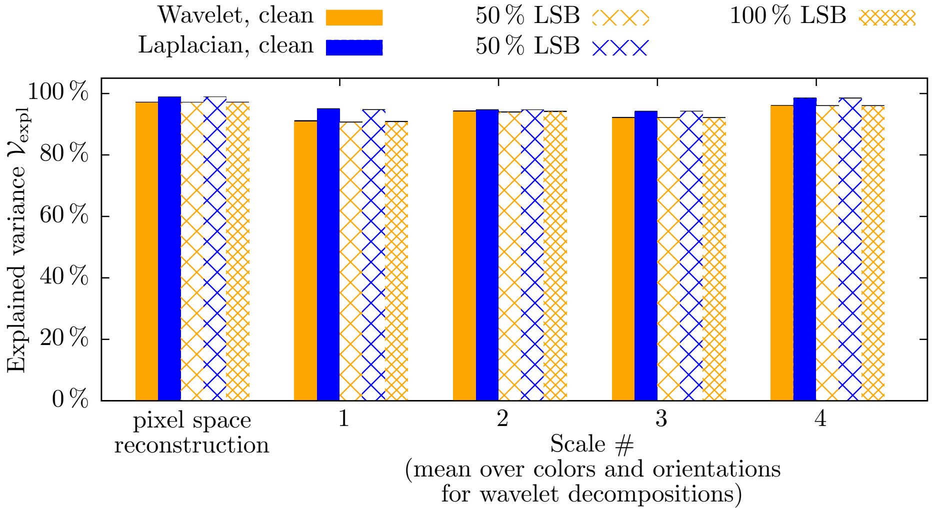


Fig. 4. Just as Fig. 3, but for color images and different embedding rates. The strong correlations between the RGB color channels improve the prediction, thus differences between the methods are considerably smaller than in grayscale images.

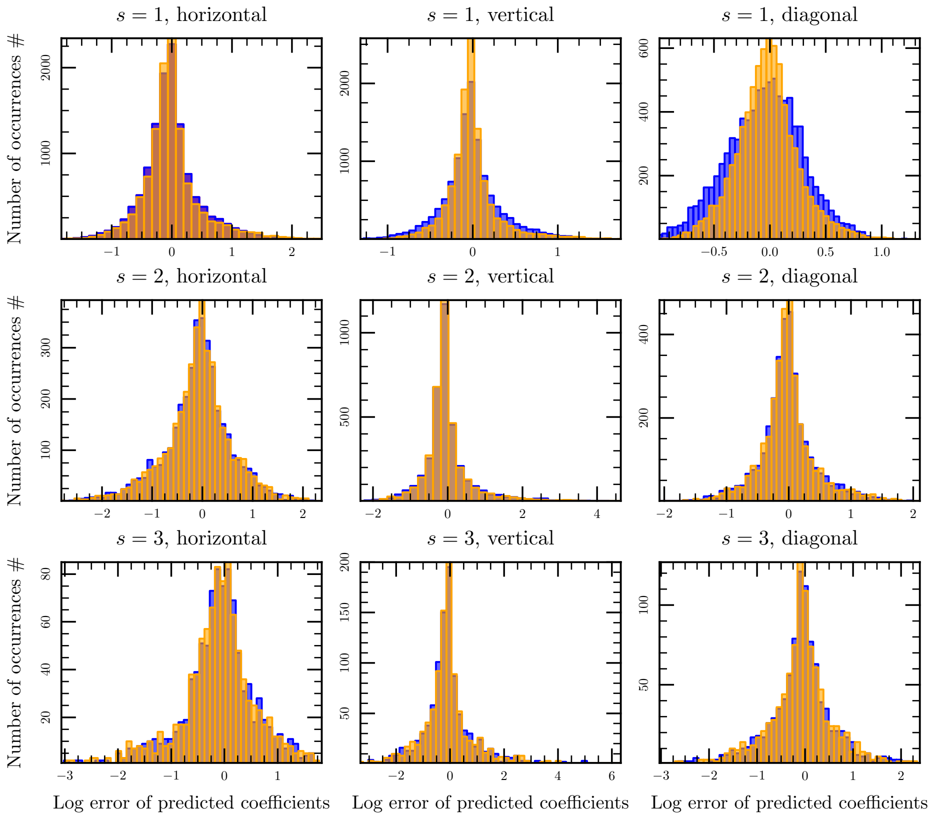


Fig. 5. Log error for each scale and orientation of a wavelet pyramid for the green channel of a clean image (🟧) showing a church vs its stego version (🟦). Uniform random noise was embedded into the least significant bit plane, with a rate of 100 %.

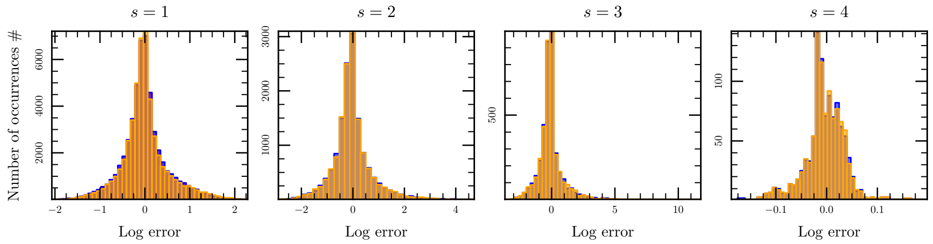


Fig. 6. Just as Fig. 5, but for the Laplacian image model.

selected the parameters of the support vector machine and its Gaussian kernel (SVM, cf. [5]) C and γ by employing a 5-fold cross-validiation scheme specifically adapted to the steganalysis scenario ([16], in preparation) on 1000 clean and 1000 stego images. Subsequently, we trained the SVM using the values of C and γ determined in the last step on 2000 images (the training data, 1000 clean and 1000 stego images), and then tested with a set of 1200 examples (600 clean and 600 stego images). We randomly divided the entire set into training and test sets and averaged over 100 splittings, which enabled us to estimate the error of the detection rate on the test set. The results showed a considerable variance for the standard test scenario of steganalysis with a fixed false positive rate of 1 %, but averaging over the 100 splittings turned out to be sufficient for finding statistically significant differences between the Laplace and wavelet basis.

We tested with two distinct normalization methods: Either each component of the vectors is scaled independently to $[0, 1]$; or only the quantile range $Q_{0.95} - Q_{0.05}$ of a component is standardized to $[0, 1]$, while clipping values above and below. We call the former standard and the latter interquantile normalization. With normalization, features of high magnitude exert a less dominant influence on the results. Additionally, for interquantile normalization, the detrimental effects of outliers are reduced.

If adapting the wavelet-based classifier such that the false positive rate falls below 1 %, then (12.6 ± 0.2) %, (30.8 ± 0.7) %, and (69.5 ± 0.9) % of the stego images can be detected for standard normalization and embedding rates of 10 %, 25 %, and 50 %, respectively. When normalizing the interquantile range of the features to $[0, 1]$, the detection rates are (8.5 ± 0.2) %, (14.4 ± 0.7) %, and (65.7 ± 1.4) %. This is a higher performance than found by Lyu & Farid in [13]. The accuracies are averaged over 100 test runs. These runs differ in that we randomly divided the entire set into training and test sets repeatedly. The full set of the results is shown graphically in Fig. 7, plus the prediction rate on ternary embeddings [10, 12] and on ±1 embeddings. ±1 embeddings conserve parity properties of the images as it is described in [12]. In this case, for embedding rates of 50 % and 25 %, (76.8 ± 1.4) % and (37.1 ± 0.6) % of the stego images can be revealed for standard normalization, and (81.9 ± 0.7) % and (20.7 ± 1.0) % for quantile normalization. The error bars denote the standard error of the mean $\overline{\sigma} = \sigma/\sqrt{n}$ over the $n = 100$ test runs, where $\sigma = \sqrt{\mathrm{Var}(x)}$ and x is the true positive rate found.

The prediction accuracies for higher embedding rates are frequently higher for standard normalization. However, for a reduced feature vector, containing only error residuals or even a subset thereof, interquantile normalization allows for higher detection rates.

In Fig. 7, Laplacian predictions are only shown for the highest embedding rate of 50 %, because their accuracies fall rapidly to values close to that achieved by random guesses: The true positive accuracies are (4.5 ± 0.1) % and (3.9 ± 0.1) %, for standard and interquantile normalization, respectively. Obviously, the detection rates of the Laplacian steganalyzer are inferior to those of the wavelet steganalyzer.

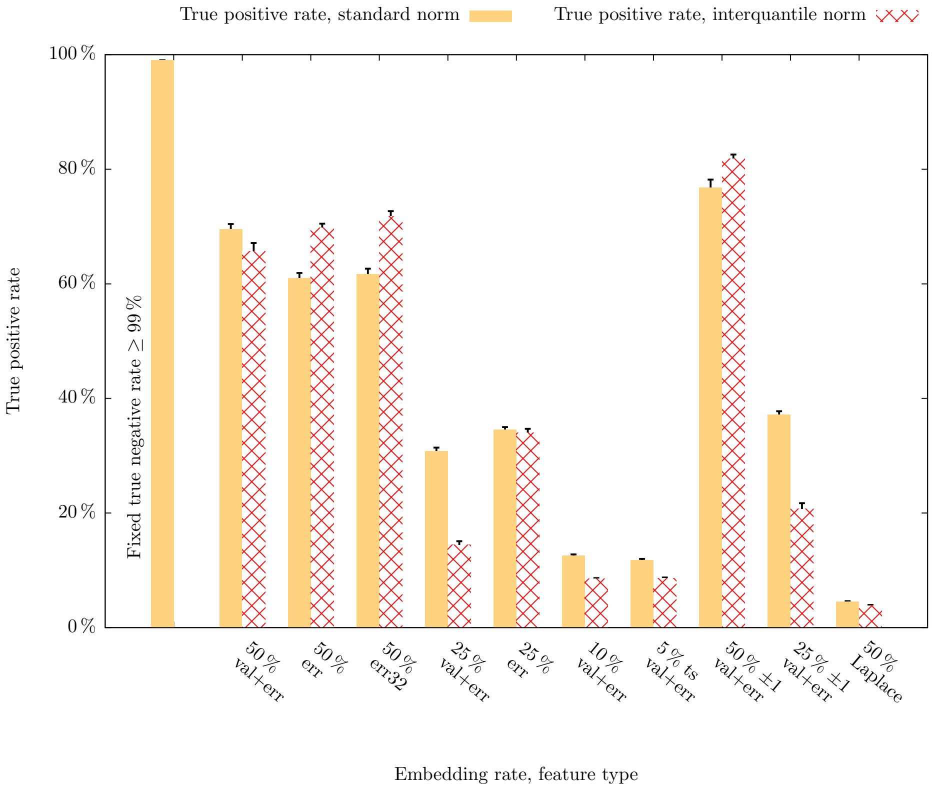


Fig. 7. Comparison of classification accuracies for four distinct embedding rates (50 %, 25 %, 10 %, 5 %), for two types of predictions (wavelet, Laplacian) and two normalization methods (interquantile, standard). The used embedding scheme is least significant bit replacement, except for “ts”, which indicates ternary, scanner-adaptive embedding, and “±1”, which indicates ±1 embedding. “err+val” denotes the usage of marginal and log error statistics, “err” denotes the utilization of log error statistics only, “err32” that of the log error statistics of the $3 \cdot 32$ moments originating from the finest scales.

4 Discussion

The hypothesis we investigate is that a better image model should yield a better detection rate. According to Figures 3 and 4, the Laplacian representation is better than the wavelet representation in terms of explained variance. However, the steganalysis experiments show that the detection performance of the wavelet model is significantly higher than that of the Laplacian so that our plausible hypothesis turned out to be wrong. As a consequence the wavelet representation must have additional properties that allow for a better steganalysis performance in spite of its inferior modeling capability.

We think that the improved discriminability in the wavelet domain can be mainly attributed to two reasons:

1. In comparison to the Laplacian, the dimensionality of the feature vectors in the wavelet representation is tripled since there are more subband statistics. It is a well-known fact (see, e. g., [9]) that the probability that two classes are separable increases with the dimensionality of their representation.
2. As described in Sect. 2, the steganalyzer uses a threshold on the coefficients before estimating the statistical moments of the subband coefficients. This turns out to be a critical step since this prevents small coefficients from influencing the feature vectors used for classification. This type of thresholding estimator is a well-known concept in signal processing where such estimators are used for denoising. We can think of the prediction step in the steganalyzer as a denoising step since we reconstruct the "true" or denoised image from the contaminated or "noisy" stego image. Here comes into play a theorem by Donoho & Johnstone [6] which states that a threshold estimator has a higher denoising performance if the signal representation is sparse. Sparse in this context means a representation of a signal that needs only a few coefficients to approximate a signal with low error. Although we did not formally test this on our data, it is a common observation that wavelets are a sparser representation of natural images when compared to the Laplacian pyramid. As a consequence, we can expect a more accurate estimation of the residual in the wavelet representation leading to better estimates of the subband statistics which finally results in a higher classification performance.

The advantage of a high-dimensional representation can be demonstrated in a simple experiment: On our test dataset of the 1600 color image pairs in uncompressed format with a 50 % LSB embedding (see Sect. 3.2), we first computed the model predictions in the Laplacian domain and transformed the predicted images back into their pixel representation. In the next step, both original image and stego image were wavelet-transformed and subjected to the same analysis as before. In this way, the modeling step took place in the Laplacian domain, whereas the classification step was based on feature vectors with the three-fold dimensionality of the wavelet domain. As a result, detection performance of this hybrid steganalyzer improved considerably from a true positive rate of (4.5 ± 0.1) % of the pure Laplacian steganalyzer to (37.3 ± 1.0)%, although the performance of the pure wavelet steganalyzer of (69.5±0.9) % could not be achieved. The results are given at a fixed true negative rate of 99 %.

A second experiment highlights the critical influence of the thresholding step: Disabling the thresholding process in the wavelet steganalyzer reduces the detection performance from (69.5 ± 0.9) % to (27.1 ± 0.6) %. This demonstrates that wavelet thresholding plays a role of similar importance to the higher dimensionality of the resulting feature vectors.

Finally one might ask why the hybrid steganalyzer from the first experiment did not reach the performance of the pure wavelet steganalyzer. The reason for this can be seen in a third experiment where we analyzed the explained variance in the wavelet domain of both the Laplacian model predictions and the wavelet model predictions (cf. Table 1). The results show that, although the Laplacian model is more accurate in its own domain, it does not reach the accuracy of the wavelet model in the wavelet domain. In our opinion, this accounts for the observed performance difference between the hybrid and the pure wavelet steganalyzer.

Table 1. Mean over orientations for wavelet decompositions in % explained variance of the hybrid and the pure wavelet steganalyzer.

	Clean images (wavelet)	Clean images (hybrid)	Stego images (wavelet)	Stego images (hybrid)
Pyramid level 1	89.6 %	88.7 %	88.5 %	87.9 %
Pyramid level 2	91.8 %	82.5 %	91.2 %	81.3 %
Pyramid level 3	91.9 %	94.3 %	91.4 %	94.2 %

5 Conclusion

In this study we analyzed the relationship between image modeling and detection performance in a universal Lyu & Farid type steganalyzer. Our results show that a high performance in image modeling does not directly transfer to a higher steganalysis performance. Although the Laplacian representation leads to a better image model, it shows an inferior detection performance. From the steganalysis point of view, the characteristics of the wavelet representation (sparse representation and higher dimensionality of the feature vector) turned out to be more important as it evidently allows the final classification stage to discriminate between the resulting feature vectors more easily. Furthermore, it seems important to stay in the same transformation space in all steganalysis steps from image modeling, thresholding, computation of feature vectors to classification. This study shows a connection between sparsity of the image basis, denoising and steganalysis performance. In future work, we plan to make this link more explicit.

6 Acknowledgments

P. H. D. Le was supported by the grant "Detection of Steganography in Images with Statistical Models" of the German Federal Ministry of Research and Education, and partially—together with M. O. Franz—by the grant "Communication and Automation" of the Baden-Württemberg Foundation. This work was also partially supported by the German National Academic Foundation. We would like to thank M. Messmer, O. Schönemann, Medav GmbH (Ilmenau) and J. Keppler for their help in the project, and H. Schwigon of the German Federal Office for Information Security for providing us useful data and discussions.

References

1. Adelson, E.H., Burt, P.J.: Image data compression with the Laplacian pyramid. In: Proceedings of the 1981 Conference on Pattern Recognition and Information Processing. pp. 218–223. IEEE Computer Society Press (1981)
2. Avcibas, I., Memon, N.D., Sankur, B.: Steganalysis using image quality metrics. IEEE Transactions on Image Processing 12(2), 221–229 (February 2003)
3. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover Publications, New York, NY, USA (June 1966)
4. Buccigrossi, R.W., Simoncelli, E.P.: Image compression via joint statistical characterization in the wavelet domain. IEEE Transactions on Image Processing 8(12), 1688–1701 (December 1999)
5. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines (2001), software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`
6. Donoho, D., Johnstone, I., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455 (1993)
7. Geisser, S., Eddy, W.F.: A predictive approach to model selection. Journal of the American Statistical Association 74(365), 153–160 (March 1979)
8. Goljan, M., Fridrich, J.J., Holotyak, T.: New blind steganalysis and its implications. Security, Steganography, and Watermarking of Multimedia Contents VIII 6072(1), 1–13 (February 2006)
9. Haykin, S.: Neural Networks (2nd Edition). Prentice-Hall, Upper Saddle River, NJ, USA (1999)
10. Holotyak, T., Fridrich, J.J., Soukal, D.: Stochastic approach to secret message length estimation in ±k embedding steganography. In: Delp, E.J., Wong, P.W. (eds.) Security, Steganography, and Watermarking of Multimedia Contents. Proceedings of SPIE, vol. 5681, pp. 673–684. International Society for Optical Engineering, SPIE, San Jose, CA, USA (2005)
11. Holotyak, T., Fridrich, J.J., Voloshynovskiy, S.: Blind statistical steganalysis of additive steganography using wavelet higher order statistics. In: Dittmann, J., Katzenbeisser, S., Uhl, A. (eds.) Communications and Multimedia Security. Lecture Notes in Computer Science, vol. 3677, pp. 273–274. Springer-Verlag, Berlin, Germany (September 2005)
12. Ker, A.D.: Improved detection of LSB steganography in grayscale images. In: Fridrich, J. (ed.) Information Hiding. Lecture Notes in Computer Science, vol. 3200, pp. 97–115. Springer-Verlag, Berlin, Germany (December 2004)
13. Lyu, S., Farid, H.: Steganalysis using higher-order image statistics. IEEE Transactions on Information Forensics and Security 1(1), 111–119 (March 2006)

14. Rasmussen, C.E.: Gaussian processes in machine learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) Advanced Lectures on Machine Learning. Lecture Notes in Computer Science, vol. 3176, pp. 63–71. Springer-Verlag, Berlin, Germany (October 2004)
15. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA (January 2006)
16. Schwamberger, V., Franz, M.O.: Simple algorithmic modifications for improving blind steganalysis performance (2010), in preparation
17. Schölkopf, B., Smola, A.J.: Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA (2002)
18. Simoncelli, E.P., Adelson, E.H.: Subband transforms. In: Woods, J.W. (ed.) Subband Image Coding. Kluwer Academic Publishers, Norwell, MA, USA (1990)