

NOISE VARIANCE ESTIMATION IN NONLOCAL TRANSFORM DOMAIN

Aram Danielyan and Alessandro Foi

Department of Signal Processing, Tampere University of Technology
P.O. Box 553, 33101, Tampere, Finland
web: <http://www.cs.tut.fi/~foi> email: firstname.lastname@tut.fi

ABSTRACT

We consider the estimation of the variance of an additive white Gaussian noise corrupting an image.

In the proposed approach, we exploit the nonlocal self-similarity of images to achieve an improved separation of noise and signal. In particular, we utilize the same adaptive 3-D transform decomposition used in the BM3D (block-matching and 3-D filtering) denoising algorithm, where mutually similar blocks are stacked together and jointly processed. An adaptive-size portion of the high-frequency ends of the 3-D transform spectra is retained and used as input sample for a robust median estimator of the absolute deviation.

Experimental analysis demonstrate a state-of-the-art accuracy of the proposed approach.

1. INTRODUCTION

The problem of estimating the statistics of the noise corrupting a digital image is a basic but fundamental step in many image processing applications. For instance, in image denoising, the amount of smoothing is adjusted based on the strength of the noise (e.g., [2],[3],[4],[6]). Similarly, when compressing noisy images, the optimal quantization steps are chosen based on the level of the noise, so that only the useful part of the data is coded [13],[15].

Roughly speaking, algorithms for estimating of the noise characteristics rely on the ability of extracting some features from the image which are affected mainly or exclusively by the noise and not by the image. The local smoothness or fast spectral decay of natural images are typical hypotheses to make the estimation possible, e.g., by measuring the noise statistics on some homogenous patches [1] or on the highest-frequency portion of the image spectrum [4],[7].

Highly structured, detailed, and textured images are more difficult to process, since part of the underlying true image often contributes to the measured features, leading to an overestimate of the noise statistics. Such overestimation spoils the precision of the following processing stages: particularly in denoising, it leads to oversmoothing of the image details.

This work was supported by the Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, project no. 118312, Finland Distinguished Professor Programme 2007-2010, and project no. 129118, Postdoctoral Researcher's Project 2009-2011) and by Tampere Graduate School in Information Science and Engineering (TISE).

In this paper, we consider the estimation of the variance of an additive white Gaussian noise (AWGN) corrupting an image. This is the simplest and most widely used model describing noise in digital images. As such, a number of different methods for estimating this variance has been proposed. The most successful ones are based on spectral decompositions of the image through a normalized decorrelating transform such as wavelets or discrete cosine transform (DCT). The noise characteristics are preserved by this operation while at the same time the energy of the underlying image is compacted into few coefficients well localized within the low-frequency portion of the spectrum. Sample statistics taken on the high-frequency end of the spectrum produce an estimate of the variance of the initial corrupting noise. The success of this approach depends on both 1) how well the transform is able to compact or sparsify the signal and 2) how insensitive the sample statistic is with respect to leakage of the original image into the samples (i.e. outliers). Therefore, improvement in estimation of the noise variance comes from utilizing better sparsifying transforms coupled by robust statistics (most notably median or nonlinear histogram fit). One of the most popular algorithms is based the median absolute deviation (MAD) of the wavelet detail coefficients of z [9],[4]. Often the image is first segmented into regions of different smoothness and then the statistics are computed only from the data belonging to the most homogeneous regions [15],[1].

In the proposed approach, we exploit the nonlocal self-similarity of images to improve the sparsification and thus achieve an improved separation of noise and signal. In particular, we utilize the same adaptive 3-D transform decomposition used in the BM3D (block-matching and 3-D filtering) denoising algorithm [2], where mutually similar blocks are stacked together into 3-D array structures called *groups* and jointly processed. An adaptive-size portion of the high-frequency corners of the 3-D transform spectra of each group is retained and used as input samples for a robust median estimator of the absolute deviation. The size of the sample portions is determined aiming at achieving an optimal bias-variance trade-off in the estimation of the median absolute deviation. This adaptation is realized through the Intersection of Confidence Intervals (ICI) rule [11],[8].

The rest of the paper is organized as follows. In the next section we formally introduce the adopted observation model and notation. A detailed description of the proposed algorithm is given in Section 3. Simulation experiments,

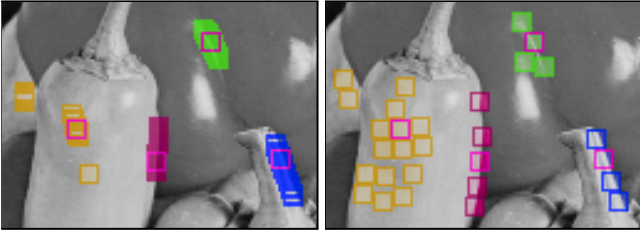


Figure 1: Grouping with overlapping vs. non-overlapping blocks.

demonstrating state-of-the-art performance of the algorithm, are presented in Section 4. Section 5 concludes the paper with some remarks about possible generalization of the approach.

2. OBSERVATION MODEL

Let z be an observed noisy image corrupted by some independent and identically distributed (i.i.d.) AWGN η with mean zero and variance σ^2 . More precisely,

$$z(x) = y(x) + \eta(x), \quad x \in X \subset \mathbb{Z}^2, \quad (1)$$

where $y : X \rightarrow Y \subseteq \mathbb{R}$ is a deterministic unknown original image and $\eta(\cdot) \sim \mathcal{N}(0, \sigma^2)$. Here, X and Y denote the image domain (pixel coordinates) and codomain (range), respectively. We assume $Y = [0, 255]$.

Our goal is to estimate the variance σ^2 from z . Nevertheless, for scaling reasons, we shall always indicate the results in terms of the estimate $\hat{\sigma}$ of the standard-deviation σ .

3. ALGORITHM

The proposed variance-estimation algorithm can be divided into two main parts.

3.1 Part 1: Blockwise and groupwise operations

3.1.1 Overlap-free grouping and 3-D transform

First we process the image in sliding-block manner: for each block Z_{x_R} in the image (here x_R stands for the coordinate of the top-left pixel of the block), we search for multiple other blocks similar to it. We call *group* the 3-D array \mathbf{Z}_{x_R} formed by stacking together these similar blocks. In this way, to each block in the image we associate a group. As in [2], we use square blocks of fixed size $N_1 \times N_1$ and group together up to N_2 similar blocks, including Z_{x_R} itself. Thus, each group has size $N_1 \times N_1 \times |S_{x_R}|$, where $|S_{x_R}|$ denotes the number of blocks in the group. We have $1 \leq |S_{x_R}| \leq N_2$ and the actual value of $|S_{x_R}|$ depends on the availability in the image of blocks sufficiently similar to Z_{x_R} . In our current implementation we set $N_1 = 8$ and $N_2 = 16$.

Groups are characterized by both intra-block correlation (between the pixels of each grouped block) and inter-block correlation (between the corresponding pixels of different blocks). The former follows from statistical properties of natural images, while the latter is a direct consequence of grouping together similar blocks. Therefore, when we apply a 3-D orthonormal decorrelating transform \mathcal{T}_{3D} on a group

we are effectively decorrelating the original signal, compacting it into few coefficients of the spectrum $\mathcal{T}_{3D}(\mathbf{Z}_{x_R})$. In this work, \mathcal{T}_{3D} is a 3-D separable discrete cosine transform (DCT).

Unlike in the grouping used in BM3D denoising, here we demand that grouped blocks are mutually *nonoverlapping*. This ensures that the noise is independent and that the 3-D transform applied on the group is equivalent to an orthonormal transform applied on the union of all pixels which belong to the grouped blocks. Thus, we have that the group spectrum is corrupted by some Gaussian zero-mean additive noise with standard deviation equal to σ . If overlapping blocks were stacked together into a group, the noise would be partly correlated, eventually leading to a severe underestimate of the noise standard deviation. Pragmatically, to construct an overlap-free group, we stack blocks one after the other, the most similar blocks first, discarding blocks that overlap with any of those already stacked. An illustration of the overlap-free grouping is given in Figure 1.

To prevent noise patterns from influencing the grouping, the similarity between blocks is evaluated by comparing corresponding blocks extracted from a denoised estimate \hat{y} of y . The ℓ^2 -norm of the difference between blocks is used for this comparison. In particular, in our current implementation, this \hat{y} is obtained from z using the standard BM3D filter, where as input standard-deviation parameter we take a rather large value, $4 \times \hat{\sigma}_{MADW}$, $\hat{\sigma}_{MADW}$ being the estimate of σ produced by the MAD of the wavelet detail coefficients of z [4]. While this results in obvious and detrimental oversmoothing of the image, it effectively remove possible structures of the noise realization. If one were to match directly on z , repeated noise patterns would be matched together and end up in contributing to the lower-frequency portion of the group spectrum, which also results in underestimation of the noise standard deviation.

3.1.2 Coefficient analysis

The transform coefficients from all groups can be classified according to their position in the group spectrum and to the number of blocks in the group to which they belong. For example, within a group, it is rather natural to use a zig-zag scan like the one illustrated in Figure 2. The idea is to sort all coefficients according to the expected energy of the underlying true signal. In particular, the coefficient in the highest-frequency corner of a group spectrum is the one that should be the least affected by the underlying image y and therefore it appears as ideal candidate for being included in a sample used to estimate the noise variance. This is particularly true for groups that include many blocks, as these groups typically provide the most effective sparsification of the underlying image.

However, large textures or structures in the image might also be captured into large groups. It is not evident whether or not such large groups should be preferred to smaller groups that correspond to a smooth region. In other words, while a zig-zag scan can be useful for sorting coefficients

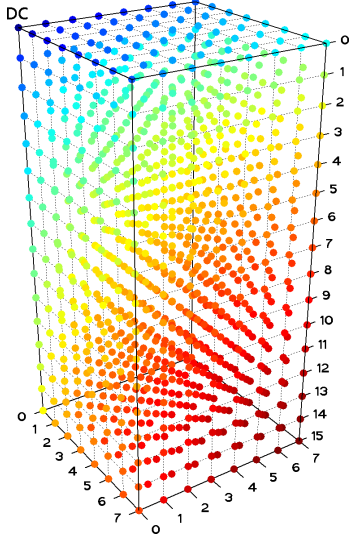


Figure 2: Ordering of the coefficients in the spectrum of a group of size $8 \times 8 \times 16$. The color of each spot represents the position of the corresponding coefficient in the zig-zag scan, starting from dark red (high-frequency corner) and ending with dark blue (DC term).

within the same group, it is difficult to decide an a-priori ordering between coefficients coming from different groups.

While it is not possible to estimate the unknown signal energy from the single noisy coefficient itself, a reasonable upper-bound of this quantity can be obtained from the average of the magnitude of few of the subsequent coefficients along the group zig-zag scan. This average is an upper-bound because it includes the energy of the noise and because the energy of subsequent coefficients in the scan is expected to grow. We use it as a surrogate for the actual energy.

In this pragmatic way, we can sort all coefficients from all groups in the highly redundant spectral representation of BM3D.

In practice, it is sufficient to collect only a tiny portion of the group spectrum: in our implementation, for each group we collect only its first 6 transform coefficients following the zig-zag scan and use up to 32 coefficients to compute the energy estimates. Thus, there is also no need to compute the full spectrum of the groups.

We denote the collection of the sorted coefficients as $\Phi_{\text{ALL}} = \{\varphi_1, \varphi_2, \dots\}$ and the expected energy (upper bound) of each coefficient as $e_j = E\left\{|\varphi_j|^2\right\}$, $j = 1, 2, \dots$

3.2 Part 2: Adaptive statistics

Once the transform coefficients from all groups have been collected and sorted, we can proceed with the calculation of the sample statistics.

3.2.1 Median of absolute deviation with varying sample size

As a robust estimate of the standard deviation, we use the conventional median of the absolute deviations (MAD) [9],[10],[14]. Given a sample Φ containing $|\Phi|$ coefficients

φ_k , $k = 1, \dots, |\Phi|$, it is defined as

$$\text{mad}\{\Phi\} = \frac{1}{0.6745} \text{median}_{\varphi_k \in \Phi} \{|\varphi_k|\}. \quad (2)$$

The standard deviation of the estimator (2) can be approximated as [17],[5]

$$\text{std}\{\text{mad}\{\Phi\}\} = \sigma \sqrt{\frac{1.35}{|\Phi| + 1.5}}, \quad (3)$$

with the distribution of errors approaching a normal. We then construct a nested sequence of samples $\{\Phi_l\}_l$, $\{\varphi_1\} \subset \dots \subset \Phi_l \subset \Phi_{l+1} \subset \dots \subset \Phi_{\text{ALL}}$, where

$$\Phi_l = \bigcup_{j=1}^{h(l)} \varphi_j,$$

and $h(l)$ is a monotonically increasing sequence of indices. The sequence of the estimates $\hat{\sigma}_l = \text{mad}\{\Phi_l\}$ satisfies two basic conditions. Because $|\Phi_j| \leq |\Phi_{j+1}|$, the estimation standard-deviation (3) is decreasing. Likewise, because $e_j \leq e_k$ for all $j \leq h(l) \leq k$, we also have that the expected estimation bias is increasing. We thus face a bias-variance trade-off scenario, where the intent is to determine a value of l for which the estimation mean squared error is minimized.

3.2.2 Adaptive sample-size selection using ICI

The Intersection of Confidence Intervals (ICI) rule [8, 11] is the criterion used to adaptively select a value of l , and hence the size of the sample used for computing the estimate of the standard-deviation (2). We implement the ICI rule as follows.

Consider the intersection of confidence intervals $\mathcal{I}_m = \bigcap_{l=l_1}^m \mathcal{D}_l$, where the confidence intervals \mathcal{D}_l are defined as

$$\begin{aligned} \mathcal{D}_l &= \\ &= \left[\text{mad}\{\Phi_l\} \left(1 - \Gamma \sqrt{\frac{1.35}{|\Phi_l| + 1.5}}\right), \text{mad}\{\Phi_l\} \left(1 + \Gamma \sqrt{\frac{1.35}{|\Phi_l| + 1.5}}\right) \right], \end{aligned}$$

$\Gamma > 0$ is a threshold parameter, and $l_1 \geq 1$ is a positive index. Let m^+ be the largest of the indexes m for which \mathcal{I}_m is non-empty, i.e. $\mathcal{I}_{m^+} \neq \emptyset$ and $\mathcal{I}_{m^++1} = \emptyset$. The adaptive sample Φ^+ is defined as $\Phi^+ = \Phi_{m^+}$, and thus the adaptive standard-deviation estimate is

$$\hat{\sigma} = \text{mad}\{\Phi^+\}.$$

The index $l_1 \geq 1$ is used to ensure that the standard-deviation estimate scaling the width of the first confidence interval $\mathcal{D}_1 = \mathcal{I}_1$ is reliable enough. We take l_1 as the smallest index l for which $|\Phi_l| \geq 500$. The threshold parameter is set as $\Gamma = 1.5$.

4. EXPERIMENTS

We present estimation results computed over a database of 25 color images (reference images of TID2008 database [16]) corrupted by noise with standard deviation $\sigma = 5$ and $\sigma = 10$. Each R, G, B component is treated as a separate image. This is a rather challenging database, as many of the images are highly textured. As reference methods, we consider the MAD on the wavelet detail coefficients [4] (with verti-

cal and horizontal Daubechies kernels of length 6) and the DCT-based method [15], which can be considered as the current state-of-the-art. Exactly the same noise realizations have been used for the three methods.

The results are summarized in Figure 3 and Figure 4. We can see that the proposed method has an accuracy which is comparable with the method [15]. Both methods clearly outperform the simpler estimator [4], which because of the textures is largely overestimating the correct value of σ .

5. CONCLUSIONS

The very good variance-estimation results achieved by the proposed approach are in line with the excellent denoising performance of the BM3D algorithm. This confirms the effectiveness of non-local transforms as a powerful tool for diverse image processing problems. We wish to remark that the reported results are preliminary, as no parameter optimization had been carried out for the algorithm.

In this work, we mainly emphasized the adaptation with respect to the transform, given both by the grouping and by the sample-size selection (which is enabled by the specific transform decomposition), rather than on the sample statistics estimator, for which we relied on the simple MAD. Better estimators could have been used as well, provided a model for their standard-deviation to be inserted in the ICI.

Although here we considered solely the basic AWGN model, the same principles can be applied to more realistic and practical noise models. In particular, our non-local transform decomposition can be embedded within the estimation algorithm for heteroskedastic observations [7], which is relevant for modeling dealing with raw data from digital imaging sensors. Let us however mention that, when facing heteroskedasticity, to avoid additional bias the block sizes cannot be arbitrarily large. Therefore, one may need to use larger blocks for determining the similarities than those eventually stacked in the group. This is a stratagem shared by many non-local filtering algorithms (see e.g., [12],[3]).

ACKNOWLEDGMENTS

We thank N. Ponomarenko and V. Lukin for kindly providing the experimental results for their method [15].

REFERENCES

- [1] Abramov, S.K., V.V. Lukin, B. Vozel, K. Chehdi, and J. Astola, "Segmentation-based method for blind evaluation of noise variance in images", *J. Appl. Remote Sens.*, vol. 2, no. 1, Aug. 2008.
- [2] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering", *IEEE Trans. Image Process.*, vol. 16, no. 8, Aug. 2007.
- [3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D Image Denoising with Shape-Adaptive Principal Component Analysis", *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*, Saint-Malo, France, April 2009.
- [4] Donoho, D.L., and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.
- [5] Foi, A., *Pointwise Shape-Adaptive DCT Image Filtering and Signal-Dependent Noise Estimation*, D.Sc.Tech. Thesis, Institute of Signal Processing, Tampere University of Technology, Publication 710, Dec. 2007.
- [6] Foi, A., "Clipped noisy images: heteroskedastic modeling and practical denoising", *Signal Processing*, vol. 89, no. 12, pp. 2609-2629, Dec. 2009. doi:10.1016/j.sigpro.2009.04.035
- [7] Foi, A., M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data", *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737-1754, Oct. 2008.
- [8] Goldenshluger, A., and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Math. Meth. Statistics*, vol. 6, pp. 135-170, 1997.
- [9] Hampel, F.R., "The influence curve and its role in robust estimation", *Journal of American Statistical Association*, 62, pp. 1179-1186, 1974.
- [10] Huber, P.J., *Robust statistics*, John Wiley & Sons Inc., 1981.
- [11] Katkovnik, V., "A new method for varying adaptive bandwidth selection", *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567-2571, 1999.
- [12] Katkovnik, V., A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising", to appear in *Int. J. Computer Vision*, 2009. doi:10.1007/s11263-009-0272-7
- [13] Lukin, V.V., M.S. Zriakhov, N.N. Ponomarenko, and A. Kaarna, "An automatic approach to lossy compression of images corrupted by Poisson noise", *Proc. Microwaves, Radar and Remote Sensing Symposium, 2008, MRRS 2008*, pp. 139-142, Kiev, Ukraine, Sept. 2008.
- [14] Mosteller, F., and J.W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*, Addison Wesley, 1997.
- [15] Ponomarenko, N.N., V.V. Lukin, M.S. Zriakhov, A. Kaarna, and J. Astola, "An automatic approach to lossy compression of AVIRIS images", *Proc. IEEE Int. Geoscience and Remote Sensing Symposium 2007, IGARSS 2007*, pp. 472-475, July 2007.
- [16] Ponomarenko, N., M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color Image Database for Evaluation of Image Quality Metrics", *Proc. Int. Workshop on Multimedia Signal Processing, MMSP 2008*, Cairns, Australia, pp. 403-408, Oct. 2008.
- [17] Rousseeuw, P.J., and C. Croux, "Alternatives to the Median Absolute Deviation", *J. American Statistical Assoc.*, vol. 88, no. 424. pp. 1273-1283, Dec. 1993.

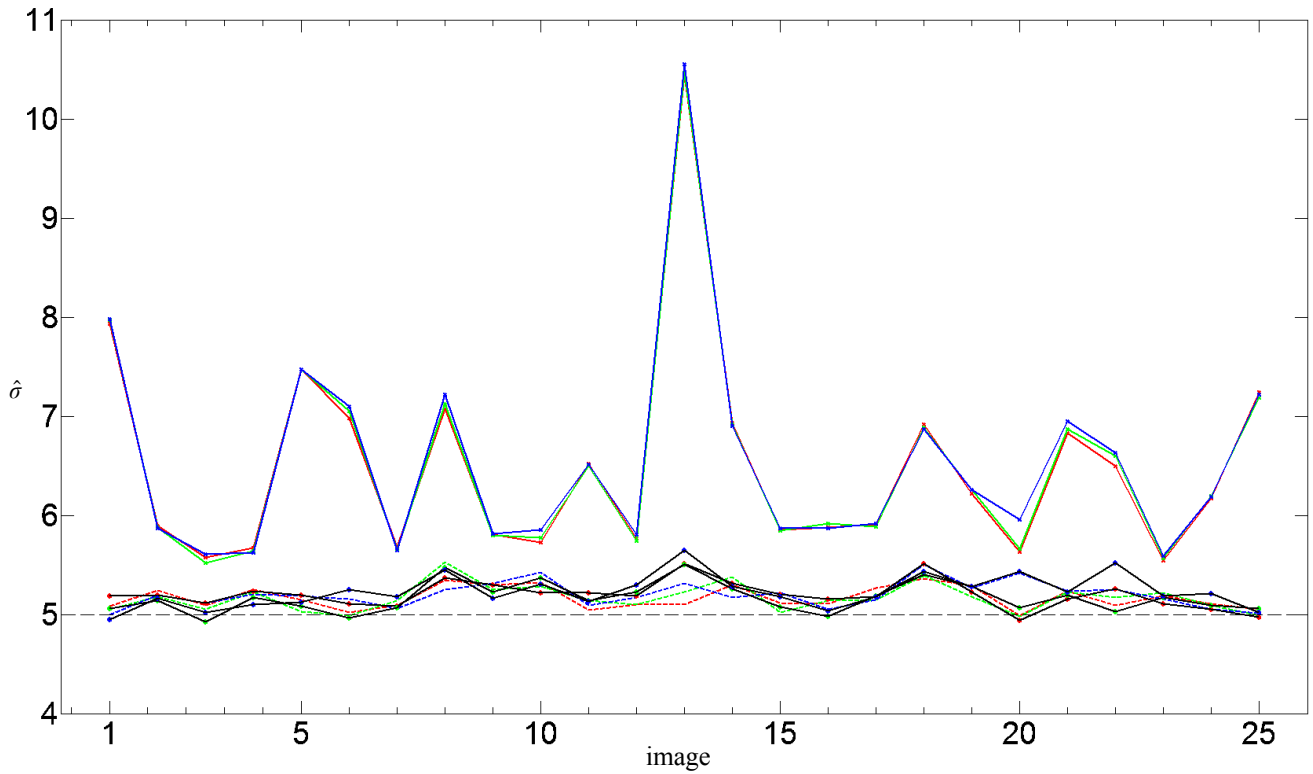


Figure 3: Estimation of the standard deviation, $\sigma = 5$. Solid colored lines: MAD on wavelet detail coefficients [4]. Dashed colored lines: block-DCT estimator [15]. Solid black lines: proposed method. Colors in the plot denote the estimation on the different R, G, or B components.

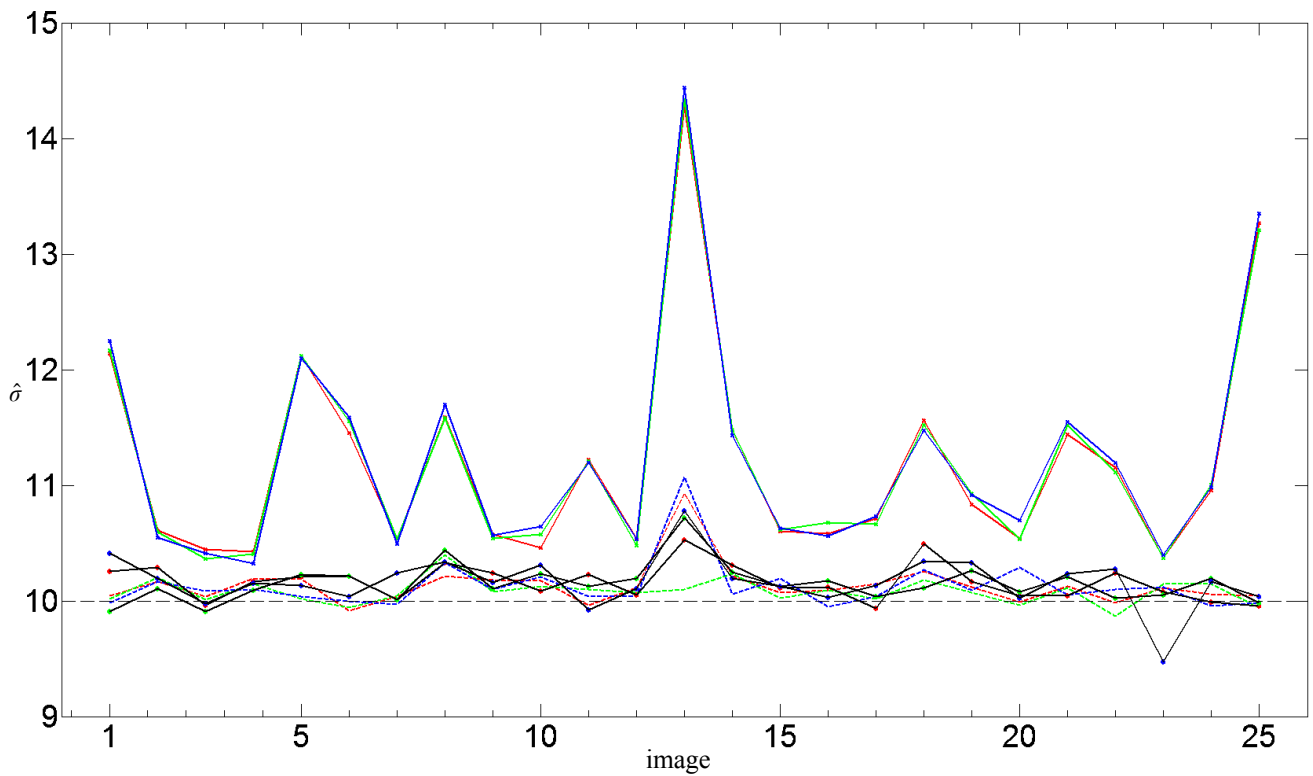


Figure 4: Estimation of the standard deviation, $\sigma = 10$. Solid colored lines: MAD on wavelet detail coefficients [4]. Dashed colored lines: block-DCT estimator [15]. Solid black lines: proposed method. Colors in the plot denote the estimation on the different R, G, or B components.